University of Alberta

Three Essays on Empirical Market Microstructure

by

Rahul Ravi ©

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Finance

Faculty of Business

Edmonton, Alberta
Fall 2007

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-33049-4
Our file Notre référence
ISBN: 978-0-494-33049-4

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Dedication

To my beloved family:

My wife Rajshree, for putting up with our "part time" relationship; My father Ravi, who has been an example and a mentor throughout my life; my sisters Guddi and Shefali for their support and love;

and especially to my mother Lakshmi, who brought me into this world and taught me the values of truth, endurance and patience.

Abstract

The first essay of this dissertation examines the time-series evolution and the cross-sectional variation of information asymmetry, as measured by the adverse selection cost of trading in the IPO aftermarket. We find that information asymmetry is lowest immediately post-IPO and increases monotonically in the first 8 to 12 weeks post IPO. Order imbalance variability and the fraction of small trades (proxies for the extent of uninformed trading), as well as return volatility (a proxy for information arrival) emerge as the key determinants of information asymmetry.

The second essay investigates the association between 15-minute order-flow variability and the adverse selection cost of trading, stock returns, and trading volume. We find that order-flow variability is positively associated with various proxies for divergence in opinions. Our analysis also suggests that periods of high order-flow variability for a stock are likely to be followed by periods of lower returns, lower spreads, and higher volume. We find strong evidence for the co-movement in order-flow variability as well as in the adverse selection cost of trading and inventory carrying costs.

The third essay explores the relation between firm opacity and information asymmetry, as measured by the adverse selection cost of trading. Existing studies interpret level of opacity as a measure of firm-to-investor information asymmetry. Adverse selection cost of trading is a measure of information asymmetry between investors. We find evidence for a significant non-monotonic relation between opacity and the level of adverse selection cost. The adverse selection cost of trading increases, and then declines, as transparent firms become increasingly opaque.

Acknowledgement

First, I thank my advisor Dr. Aditya Kaul, for his continuous support in the Ph.D. program. He has been always there to listen and to give advice. He is responsible for involving me in market microstructure research. He taught me how to ask questions and express my ideas. He showed me how to approach a research problem and the need to be persistent to accomplish any goal.

I thank Dr. Mark Huson for being such a nice mentor to me. He showed me how to work hard simply by doing it, himself. We talked a lot and Mark was always a very patient listener. I thank Dr. Vikas Mehrotra for being a friend, a guide and a mentor throughout the PhD process. Both, Mark's and Vikas' incisive comments have been very helpful in shaping and focusing my research.

I would like to thank my external committee member Dr. Marc Lipson for carefully reading through my thesis and providing insightful comments and suggestions. I also thank my committee members Dr. Constance Smith and Dr. Barry Scholnick for their comments and helpful suggestions.

My thanks also go to Dr. David Cooper for his constant support and his faith in my abilities. He has rescued me from various red tape crises. Without his encouragement and guidance, I could not have finished this dissertation.

Completing the research underlying this dissertation required very difficult levels of computer programming. I would like to thank Mary Gillespie for improving my programming skills and for always being there for me during times of database crises.

I thank the support staff at the School of business for all their help. Kathy Harvey, Jeanette Gosine, Keltie Tolmie and Louise Hebert: thank you for all your support. Without your help completing this program would have been much more diffcult.

I thank my friends Kathy Fogel, Gloria Tian, Saif Ullah Warraich and Igor Semenenko among several others whose presence provided me with emotional sanity and whose helpful suggestions helped me complete this dissertation. I especially thank my friend Larry Hunter. Our hiking and camping trips helped me reduce stress.

My thanks go to my wife Rajshree. I am truly blessed to have such a wonderful companion. Kitta, thank you for being there and for bearing with my many years of devotion to this thesis. Those times were truly ours and now that I am done, I hope to make up to you for all those lost moments.

I thank my parent in-laws for their encouragement and support. I am greatful to them for always standing by me. Finally, I am thankful to my parents. I will forever be grateful for the values you taught me, the opportunities you gave me, and the confidence you instilled in me.

Rahul Ravi

September 26, 2007

Table of Contents

Chapter 1 1
Introduction 1
References 6
Chapter 2 7
The Secondary Market Evolution of Information Asymmetry 7
  2.1 Introduction 7
  2.2 Hypothesis Development 12
    2.2.1 Evolution of information asymmetry in the IPO aftermarket 12
    2.2.2 Determinants of aftermarket information asymmetry 15
      2.2.2.1 Variability of order-flow (σ_{OF}) 15
      2.2.2.2 Proportion of small and medium trades 16
      2.2.2.3 Return volatility 16
      2.2.2.4 Analysts following and forecast dispersion 16
      2.2.2.5 Institutional ownership 17
      2.2.2.6 Firm characteristics 17
  2.3 Empirical Methods and Construction of Variables 18
    2.3.1 Measuring information asymmetry 18
    2.3.2 Independent variables 20
    2.3.3 Model specifications 22
  2.4 Sample Selection and Sample Characteristics 24
    2.4.1 Sample selection and data 24
    2.4.2 Sample characteristics 26
  2.5 Information Asymmetry in the IPO Aftermarket 27
  2.6 The Determinants of Information Asymmetry 34
    2.6.1 Trading characteristics and information asymmetry 35
    2.6.2 Firm characteristics and information asymmetry 36
  2.7 Conclusion 37
References 39
Appendix 44
Chapter 3 60
Understanding the second moment of order-flow: Implications for the trading environment 60
  3.1 Introduction 60
  3.2 Hypotheses Development 62
    3.2.1 SIGOF and trading costs 63
    3.2.2 SIGOF and trading volume 63
    3.2.4 SIGOF and dispersion in analyst forecasts 64
    3.2.5 SIGOF and returns 65
    3.2.6 SIGOF and market-wide divergence in opinions 65
    3.2.7 Co-movement in SIGOF and liquidity 66
  3.3 Construction of Variables and Empirical Methods 66
    3.3.1 Measuring order-flow variability 67
    3.3.2 Measures of liquidity 68
    3.3.3 The adverse selection and inventory cost components of the spread 68
    3.3.4 Other variables 69
    3.3.5 Co-movement in SIGOF 72

3.4 Sample Selection and Sample Characteristics 74
3.5 Results 76
3.5.1 Exploring order-flow variability 77
3.5.2 Commonality in order-flow variability 83
3.6 Conclusion 86
References 87
Appendix 90
Chapter 4 102
Firm opacity and financial market information asymmetry 102
4.1 Introduction 102
4.2 Research Method 105
4.2.1 Box-Cox transformations 107
4.2.2 Polynomial regression 108
4.3 Measure of Adverse Selection Cost of Trading 109
4.4 Proxies for Firm-to-Investor Information Asymmetry 110
4.4.1 Proxies based on disclosure quality 111
4.4.1.1 AIMR scores 111
4.4.1.2 The S&P T&D scores 112
4.4.2 Proxies based on firm characteristics 113
4.4.2.1 Discretionary accruals 113
4.4.2.2 Firm size 114
4.4.2.3 Market-to-book ratio 114
4.4.2.4 R&D to sales ratio 115
4.4.3 Financial analyst based proxies 116
4.4.3.1 Number of analysts providing earnings forecasts 116
4.4.3.2 Analysts' forecast error 117
4.4.3.3 Coefficient of variation of analysts' forecasts 117
4.4.4 Control variables 117
4.5 Sample Selection and Sample Characteristics 120
4.6 Empirical Analysis and Results 124
4.6.1 Univariate analysis 125
4.6.2 Box-Cox transformations 128
4.6.3 Univariate regression analysis 129
4.6.4 Multivariate regression analysis 130
4.6.4.1 Market to book ratio (MB) 131
4.6.4.2 Number of analysts providing earnings forecasts (LnAnal) 132
4.6.4.3 Coefficient of variation (CV) and errors (EPE) in analysts' forecasts 133
4.6.4.4 Discretionary accruals (DAC) 134
4.7 An Alternative Test for the Non-Monotonic Relation between Inter-investor and Firm-to-Investor Information Asymmetry 135
4.8 Conclusions 137
References 139
Appendix 145

List of Tables

Table 2.1: Pair-wise correlation between measures of information asymmetry. 47

Table 2.2(A): Summary firm characteristics for NYSE and NASDAQ listed IPOs. 48

Table 2.2(B): Summary offer characteristics of the IPO firms in Table 2.1(A). 48

Table 2.3: Correlation coefficients. 49

Table 2.4: Time-series patterns in information asymmetry 50

Table 2.5: Time-series analysis of weekly information asymmetry 51

Table 2.6: Difference between information asymmetry for IPOs vs. carve-outs. 53

Table 2.7: Cross sectional determinants of weekly information asymmetry. 54

Table 2.8: Cross-sectional determinants of monthly information asymmetry. 59

Table 3.1: Distribution of firms across the sample period. 90

Table 3.2: Distribution of SIGOF by Industry. 91

Table 3.3: Time-series and cross-sectional variation in average SIGOF. 92

Table 3.4: Distribution of SIGOF 93

Table 3.5: Non-Parametric Correlation Coefficients (Spearman's rank correlation). 94

Table 3.6: Stock return, trading volume, spread and components of spread across SIGOF quintiles. 95

Table 3.7: Attributing SIGOF to firm and trading characteristics 96

Table 3.8: Attributing time-series changes in SIGOF to changes in systematic and firm specific factors. 97

Table 3.9: Market-wide commonality in SIGOF and liquidity. 98

Table 3.10: Market-wide commonality in levels of liquidity. 99

Table 3.11: The contribution of co-movement in SIGOF to co-movement in liquidity... 100

Table 3.12: Some explanations for market wide commonality in SIGOF. 101

Table 4.1: Distribution of firms across the sample period. All presented numbers are arithmetic averages 152

Table 4.2: Descriptive Statistics for NYSE-listed sample Firms (1993 to 2002) 153

Table 4.3: Pearson Correlations between various proxies of information asymmetry... 154

Table 4.4: Univariate analysis. 155

Table 4.5: Multivariate regression analysis. 156

Table 4.6: Effect of focus enhancing spin-off. 157

List of Figures

Figure 2.1(A): The evolution of post-IPO adverse selection cost of trading. 44

Figure 2.1(B): The evolution of post-IPO quoted spreads. 44

Figure 2.2: Order-flow variability. 45

Figure 2.3: The proportion of medium trades in the IPO-aftermarket. 46

Figure 4.1: Annual average level of the adverse selection cost per $100 traded. 145

Figure 4.2: Average adverse selection cost of trading, per $100 traded vs. quality of the firm's disclosure (2002 S&P disclosure ranks). 146

Figure 4.3: Observed association between λ and the various proxies of firm-to-market information asymmetry. 147

Figure 4.3(A): S&P T&D Scores. 147

Figure 4.3(B): AIMR disclosure quality rank. 147

Figure 4.3(C): Coefficient of variation of analyst forecasts. 148

Figure 4.3(D): Analyst earnings prediction error. 148

Figure 4.3(E): Performance matched discretionary accruals. 149

Figure 4.3(F): Market to book ratio. 149

Figure 4.4: The functional form between two of the firm-to-investor information asymmetry are estimated using the Box-Cox transformation technique. 150

Figure 4.4(A): Box-Cox fitted model for S&P T&D (2002) data. 150

Figure 4.4(B): Box-Cox fitted model for AIMR data. 151

Chapter 1

Introduction

The phrase "Market Microstructure" was coined by Mark Garman in 1976 as the title of his paper on trading processes. Since then, it has become the descriptive title for an area of finance that examines the economic forces affecting trades, quotes, and prices in financial markets. Madhavan (2000) defined market microstructure as the study of "the process by which investors' latent demands are ultimately translated into prices and volumes". Stoll (2003) describes it as a "field which studies the cost of trading securities and the impact of trading costs on the short-run behaviour of securities prices".

Since its inception, the demand for research in this field has grown under the influence of various factors such as market anomalies (crashes and bubbles), creation of new financial instruments, growth of new and existing forms of markets, and growth in public investments in financial markets. In more recent times, the availability of extensively detailed intra-day data and more powerful computers has given a boost to empirical research in market microstructure.

This thesis consists of three essays on empirical market microstructure. The first essay studies the evolution of information asymmetry between traders, as a publicly traded firm grows older. The second paper explores the relation between order-flow variability and divergence in opinions, in the financial market. The third paper attempts to investigate how the information asymmetry between a firm and an investor is related to the information asymmetry among investors.

The first essay explores the evolution of the information asymmetry among traders in the IPO aftermarket. We expect information asymmetry between investors to decline as the firm matures because the degree of information asymmetry about a firm's fundamentals generally decreases with age (Lang, 1991). Nevertheless, information asymmetry among investors is also affected by two additional factors. It is inversely related to the variability of uninformed traders' activities (Kyle, 1985) and directly related to the traders' abilities to process the available information into superior forecasts of firm value (for instance, see Kim and Verrecchia (1994) and Kandel and Pearson (1995)). The immediate post-IPO market is characterized by the presence of heavy uninformed trading (Aggarwal and Rivoli, 1990) and the existence of several regulatory restrictions on release of information (quiet period, penalty bids, etc.). Therefore, information asymmetry could be relatively low immediately after the IPO. As more public information arrives, information asymmetry would increase as investors see and take advantage of additional public signals.

We find that information asymmetry is lowest immediately post-IPO and increases monotonically in the first 8 to 12 weeks of the secondary market trading. Order imbalance variability and the fraction of small trades (proxies for the extent of uninformed trading) appear to generate this pattern in information asymmetry – both are at their highest levels during the first full week of trading, and progressively decay over the next 8 to 12 weeks. Our cross-sectional analysis suggests that these same two variables, as well as return volatility (a proxy for information arrival), are the key determinants of information asymmetry. Variables such as analyst following and institutional ownership are not consistently significant. Overall, our results point to information arrival and, more importantly, uninformed trading as the determinants of post-IPO information asymmetry.

The second essay studies the relation between the standard deviation of 15-minute order imbalance and stock returns, with the bid-ask spread and its components, and trading volume. We use high frequency data to first compute 15-minute order imbalance and then calculate the standard deviation of this series within each month for a sample of 3870 NYSE stocks over the period January 1993 through December 2003.

We find that, on average, a higher SIGOF leads to a lower per dollar adverse selection cost of trading, a lower inventory cost (per dollar), a lower bid-ask spread and proportional spread, lower risk-adjusted returns, and higher trading volume. Nevertheless, the negative relation between lagged order-flow variability and inventory costs is puzzling, since a more variable order-flow is the result of either a greater divergence in opinions or heavier liquidity trading, and order-flow becomes less informative about the true price and the adverse selection risk of the market maker is reduced (Kyle 1985). The positive relation between SIGOF and volume, and the negative relation between SIGOF and future returns are consistent with predictions of the divergence, from the opinion literature (Miller, 1977). We also find that SIGOF is positively associated with other proxies for divergence in opinions, such as: market capitalization, S&P 500 futures open interest, dispersion in analysts' forecasts, and the volatility of trading volume.

The second essay concludes with an investigation of order-flow co-movement. Our interest in common effects and order-flow variability is motivated by the idea that differences in opinions could be correlated across stocks. We present evidence of significant commonality in order-flow variability and find that about 83% of the stocks in our sample display varying levels of co-movement. We also find strong evidence for co-movement in order-flow variability, as well as in the adverse selection cost of trading and inventory carrying costs. Co-movement in order-flow variability appears to partially explain co-movement in liquidity and in both the adverse selection and inventory costs.

The final essay examines the relation between the information asymmetry between the firm and the investor, and the information asymmetry among investors. It is important to understand the association between the two types of information asymmetries since both affect the investment abilities of the firm. On one hand, reduction in the firm to investor information symmetry increases the availability of capital, and thus allows the firm to invest in erstwhile non-feasible, but positive, NPV projects. On the other hand, an increase in the information asymmetry among investors reduces liquidity (Kyle, 1985), which increases the cost of capital (Ahimud and Mendelson, 1988), to render some positive NPV projects non-feasible. While a firm's cost of capital is identically affected by both forms of information asymmetries, one must understand the association between the two to be able to understand the effect of its transparency and disclosure-related decisions.

We find that, after controlling for the effects of market microstructure and liquidity, a significant, non-monotonic relation exists between the firm and the investor information asymmetry, and with the information asymmetry among investors. As the level of firm-to-investor information asymmetry increases, the information asymmetry among investors rises until reaching a certain point, when it starts to decline. This result is intuitively appealing. If a firm is completely transparent, all market participants know everything about the firm and hence, the adverse selection should be zero.[1] If the firm is completely opaque, all participants are uninformed and hence, the adverse selection problem should again reduce to zero. Somewhere between the two extremes, the adverse selection cost attains its maximum.[2]

---

[1] A transparent firm has low to nil firm-to-investor information asymmetry. As the level of firm-to-investor information asymmetry declines, firms will become less transparent (or more opaque). This paper uses transparency and firm-to-investor information asymmetry synonymously. Opacity is the antonym of transparency.

[2] A caveat is in order here. For very opaque firms, the point of equilibrium inter-investor IA will be determined through the interplay of the search cost and the economic value of information. Therefore, the level of inter-investor information asymmetry might be non-zero.

References

Aggarwal, R. and P. Rivoli, 1990. Fads in the Initial Public Offering Market? *Financial Management* 19, 45–57.

Amihud, Y., and Mendelson, H., 1988. Liquidity and asset prices - financial management implications. *Financial Management*, 17(1): 5-15.

Garman, Mark B., 1976. Market Microstructure, *Journal of Financial Economics*, 3: 257-275.

Hans R. Stoll, 2003. Market Microstructure, *Handbook of The Economics of Finance*, Volume 1A: Corporate Finance, Chapter 9, pp. 555-600.

Kandel, E., and Pearson N.D., 1995. Differential interpretation of public signals and trade in speculative markets. *Journal of Political Economy*, 103(4): 831-872.

Kim, O., and Verrecchia, R.E., 1994. Market liquidity and volume around earnings announcements. *Journal of Accounting and Economics*, 17(1-2): 41-67.

Kyle, A., 1985. Continuous auctions and insider trading, *Econometrica* 53, 1315-1336.

Lang, Mark, 1991. Time-varying stock price response to earnings induced by uncertainty about the time-series process of earnings. *Journal of Accounting Research* 29, 229-257.

Madhavan, Ananth, 2000. Market Microstructure: A Survey. Available at SSRN: http://ssrn.com/abstract=218180 or DOI: 10.2139/ssrn.218180

Miller, E.M., 1977. Risk, uncertainty and divergence of opinion, *Journal of Finance* 32, 1151–1168.

Chapter 2

The Secondary Market Evolution of Information Asymmetry

2.1 Introduction

Financial economics distinguishes two broad classes of investors who differ in their motives for trade. The first group consists of informed investors, who trade because they possess information about the firm's value. The second group of investors, usually referred to as noisey, liquidity, or uninformed traders, trades for reasons unrelated to the value of the firm. The market microstructure literature argues that if markets are not strong and efficient, informed investors can profit at the expense of uninformed investors. Market-makers, presumed to be uninformed, are aware of this adverse selection problem, and adjust their bid-ask spreads to reflect the risk of information-based trading.

This paper explores the evolution of the adverse selection cost of trading as measured by Glosten and Harris (GH, 1988) in the IPO aftermarket. Since this cost results from information asymmetry between traders, we refer to it simply as information asymmetry. The paper has two objectives. First, it examines the time-series variation in information asymmetry as a function of a firm's post-IPO age. Second, it examines the determinants of the cross-sectional variation in information asymmetry as the IPO firms mature.

Our results are important for several reasons. First, information asymmetry has been shown to affect the required rate of return (e.g., Brennan and Subrahmanyam, 1996). Thus, distinct patterns in information asymmetry have the potential to show up in both a firm's financing costs and in its aftermarket performance. Second, measures of information asymmetry, such as the GH (1988), capture the permanent impact of transactions on prices. Therefore, our results shed light on the process of price formation for an extended post-IPO period. As a related matter, post-IPO trading is heavily regulated and the relaxation of regulations at clearly-delineated times potentially changes the mix of uninformed and informed traders. This study addresses the issue of whether or not significant changes occur in the adverse selection cost of trading at such times. In our cross-sectional analysis, we examine the importance of trading variables such as the proportion of small (or medium) trades, order-flow variability and return volatility, and firm characteristics such as growth opportunities, analyst following, and institutional shareholding, in explaining inter-firm variations in information asymmetry at various points after the IPO. Our analysis explores such issues as whether or not analysts play a part in leveling the playing field early in the life of a firm, and whether or not the presence of institutional shareholders exacerbates the adverse selection problem in the aftermarket.

We follow 289 NYSE and 1661 NASDAQ IPOs between 1993 and 1998 for a period of 2 to 4 years after the IPO.[3] We find that information asymmetry, computed weekly, increases for approximately 8 to 12 weeks for this sample of IPOs before it flattens out for the rest of the 2-year sample period. Kyle (1985) provides a useful framework in which to interpret this seemingly counterintuitive pattern. Kyle points to prior uncertainty about firm fundamentals and the volume of informed, relative to uninformed, trading as determining the level of information asymmetry.

---

[3] As explained later in the paper, while the frequency of analysis used in this paper is weekly, some of the cross-section analysis is carried out at a monthly frequency. While the weekly analysis requires two years of data, the monthly analysis requires the existence of four years post-IPO trading data.

Following Kyle (1985), we use the variability of order-flow (aggregated to the 15-minute level) as a proxy for uninformed trading. Order-flow variability is highest immediately post-IPO, and decays over the following 8 to 12 weeks. The correspondence between the patterns in order-flow variability and information asymmetry suggests that time variation in uninformed trading activity might explain the puzzling pattern in the post-IPO information asymmetry. We continue this investigation by examining the time-series variation in the fraction of small- and medium-sized trades. Building on Easley and O'Hara (1987), Barclay and Warner (1993) argue that informed investors will place medium-sized trades, while uninformed traders are more likely to place small trades. We study the proportion of small trades (trade size less than 500 shares) and medium trades (500 to 9900 shares) in the weeks after the IPO, and find a reduction in the fraction of small trades (equivalently, an increase in the fraction of medium trades) over the first 8 to 12 weeks.

This evidence suggests that uninformed trading is heavy, relative to informed trading, right after the IPO and then declines thereafter. This is consistent with Aggarwal and Rivoli (1990), who find that uninformed traders are attracted to IPOs because of underpricing and the resulting media attention. At least two plausible reasons may be given to explain why the extent of informed trading will be low initially and increase over time. First, informed investors might trade less aggressively immediately after the IPO, until they are able to precisely assess the extent of uninformed trading. Moreover, the end of the quiet period, 30 days following the IPO, facilitates information production (e.g., through the entry of analysts).

To study the relevance of prior uncertainty, we compare the evolution of information asymmetry for new IPOs and carve-outs. More prior information exists about carve-outs than about new IPOs. Therefore, if prior uncertainty drives the information asymmetry in the aftermarket, we should see a different level, or pattern for the evolution of information asymmetry for carve-outs vis-à-vis new IPOs. We find that the average level of information asymmetry is higher for carve-outs than for new IPOs. Nevertheless, both new IPOs and carve-outs display identical patterns in the evolution of information asymmetry. The first result suggests that investors with a superior ability to interpret information can capitalize on the greater volume of information available for carve-out firms. This interpretation is consistent with Kim and Verrecchia (1994) and Kandel and Pearson (1995). Conversely, information asymmetry is lower for IPOs because of a limited availability of information. After controlling for idiosyncratic volatility, and the extent of uninformed and informed trading, the level of information asymmetry for the carve-outs and the new IPOs is identical. This suggests that the extent of informed and uninformed trading, along with the level of information arrival, plays a key role in the evolution of information asymmetry. The similarity in the evolution of information asymmetry (especially, the initial increase) indicates that the retreat of uninformed traders exacerbates information asymmetry.

The cross-sectional analysis more closely investigates the determinants of information asymmetry. Here, we relate information asymmetry to order-flow variability, the proportion of small or medium trades, return volatility, and to firm characteristics such as size, price, book-to-market, the number of institutional shareholders and the number of analysts following the stock. In the cross-section, information asymmetry is negatively associated with the variability of order-flow and the proportion of small trades, and is positively associated with the proportion of medium trades and return volatility. It is also negatively associated with firm size. The remaining variables, including analyst coverage, are not consistently statistically significant. Overall, measures of informed and uninformed trading appear to explain some of the variations in information asymmetry, and their influence is similar throughout our event horizon.

In adding to the vast IPO literature, our interest lies in understanding the development of the trading environment for newly public firms. Our work is closest to that of the study by Corwin, Harris, and Lipson (2004) on liquidity for NYSE IPOs. They find that the NYSE-listed IPOs are characterized by unusually high limit order book depth and low bid-ask spreads at the start of trading, and trading costs (spreads) increase over time. We augment their rich analysis in several respects. While they study the bid-ask spread, we examine the component of the spread due to adverse selection. Their window of analysis extends to 30 days post-IPO, while we trace the time-series for 2 years post-IPO (our cross-sectional analysis extends to 4 years post-IPO). They examine NYSE IPOs, whereas we also study NASDAQ IPOs. Finally, they do not examine the relation between information asymmetry and firm or trading characteristics, while we make it the focal point of our analysis.

The remainder of the paper is organized as follows. Section 2.2 discusses our research questions. Section 2.3 outlines the empirical methods and defines the measures and variables used. Section 2.4 describes the sample and its characteristics. Section 2.5

studies the time-series patterns in information asymmetry. Section 2.6 examines cross-sectional variation in information asymmetry, and Section 2.7 summarizes our conclusions.

2.2 Hypothesis Development

2.2.1 Evolution of information asymmetry in the IPO aftermarket

We expect information asymmetry between investors to decline as the firm matures because the degree of information asymmetry about a firm's fundamentals generally decreases with age (Lang, 1991). Nevertheless, four additional effects need to be taken into account. Kyle (1985) provides a useful framework in which to interpret the first three effects. His measure of information asymmetry (λ) – the extent to which prices adjust in response to order-flow – is directly related to the prior uncertainty about the true value of the asset (in Kyle's notation, Σ_0) and inversely related to the variability of uninformed order-flow (σ_u^2).

First, Σ_0 is high for new IPOs because little information exists about these firms. Information is only released as part of the IPO prospectus, the road show, and the book-building process. This uncertainty should give rise to high levels of information asymmetry early in the life of an IPO, as uninformed investors learn from the trades of the informed. Departing from the Kyle model however, if investors are unable to process this limited information into superior forecasts of firm value (see, for instance, Kim and Verrecchia, (1994), and Kandel and Pearson, (1995)), information asymmetry will be relatively low immediately after the IPO. As more public information arrives, information asymmetry should increase as investors see, and take advantage of, additional public signals. Thus, we have two alternative hypotheses:

> $H_{1,1}$: Investors with superior processing ability are able to take advantage of pre-IPO information. Thus, the immediate aftermarket will be characterized by the existence of high information asymmetry between investors.

> $H_{1,2}$: Pre-IPO information is unambiguous. In the absence of new information, investors with superior processing ability are unable to derive any informational advantage. Thus, the immediate aftermarket will be characterized by the presence of a relatively small information gap between investors, and, hence, low information asymmetry.

To understand the importance of prior information, we use a sample of 171 equity carve-outs. A carve-out firm exists before its IPO, as a unit of a larger firm; thus, more information (e.g., one or more years of financial reports) is available about a carve-out firm, compared to a new IPO. This potentially alleviates information asymmetry for a carve-out firm. On the other hand, market participants with a superior ability to interpret information can capitalize on the greater information available for the carve-out firms (Kim and Verrecchia, 1994; Kandel and Pearson, 1995). Evidence to support this view of information asymmetry between investors is provided by Huson and MacKinnon (2003). If this effect dominates, the carve-out sample should exhibit a higher initial level of, and possibly a sharper decline in, information asymmetry. In either case, the patterns in information asymmetry for carve-outs and new IPOs should be different. Finally, the pattern in information asymmetry will be similar for new IPOs and carve-outs if uninformed traders are attracted to both carve-out issues and new IPOs, and the relative proportion of informed vs. uninformed traders drives information asymmetry.

> $H_{1,3}$: Patterns in information asymmetry in the early aftermarket are driven by the extent of uninformed trading. Therefore, carve-outs and new IPOs will exhibit identical patterns.

Second, information asymmetry is a declining function of the extent of uninformed trading. Studies of the post-IPO trading environment suggest that a large volume of uninformed trading (which leads to high σ_u^2) is attracted to the immediate IPO aftermarket, and that the extent of uninformed trading declines after the initial weeks of trading (Aggarwal and Rivoli, 1990; Aggarwal, Krigman and Womack, 2002). As the extent of uninformed trading declines, the market-maker can detect informed trades more easily, and information asymmetry increases. Consequently, we might see an increase in the levels of information asymmetry before it stabilizes or starts to decline.

> H_2: Declining σ_u^2 will result in increasing information asymmetry. It will stabilize as σ_u^2 stabilizes.

Third, Kyle (1985) showed that the intensity with which an informed investor trades is proportional to σ_u^2. In the absence of a trading history, σ_u^2 is initially not known to informed investors. Consequently, they might elect to trade more cautiously until they are able to estimate σ_u^2 with some precision. This suggests that the early stages of the IPO aftermarket will be characterized by uninformed order-flow, and thus, low information asymmetry. As informed traders learn the trading characteristics of the uninformed traders and increase their trading intensity, the adverse selection problem, and information asymmetry will increase.

> H_3: The volume of informed trading will be lowest immediately post-IPO. It will increase as the firm matures and more information enters the market.

Finally, the level of information asymmetry is also affected by the presence and subsequent removal of several post-IPO regulations, such as the quiet period, penalty bids, rules 144 and 171, and lockup provisions. These mechanisms are designed to keep investors or individuals who might have superior information – managers, officers and directors of the firm, pre-IPO allocation holders, or other block holders – out of the market, thereby eliminating or reducing the adverse selection risk faced by the uninformed traders. As these regulations are lifted, the proportion of informed traders should increase, leading to an increase in information asymmetry.

> H_4: Lifting of post-IPO regulations will lead to an increase in the proportion of informed trading and increase in information asymmetry.

2.2.2 Determinants of aftermarket information asymmetry

2.2.2.1 Variability of order-flow (σ_{OF})

Variability of order-flow is a measure of the extent of uninformed trading. Informed investors are better able to hide their trades when they are trading stocks with greater order-flow variability. Drawing upon the intuition discussed in the previous section, the adverse selection cost of trading should be negatively associated with order-flow variability.

2.2.2.2 Proportion of small and medium trades

As an alternative measure of the extent of informed and uninformed trading, we examine the proportion of small and medium trades. Easley and O'Hara (1987) suggest that informed investors have a preference for placing larger trades, while uninformed investors do not have similar quantity preferences. Very large (block) trades are unlikely to be motivated by information because they face large price discounts or are negotiated (so as to avoid such discounts). The results from Barclay and Warner (1993) indicate that the bulk of price-adjustment occurs in response to medium (500-9,900 shares) trades, suggesting that informed investors tend to place medium trades, while uninformed investors are likely to be concentrated in the small trade category (100 to 400 shares).

2.2.2.3 Return volatility

Return volatility is a proxy for private or public information arrival (e.g., see Ross, 1989). The arrival of private information clearly exacerbates information asymmetry. As argued above, public information can also increase information asymmetry if some investors are more adept at interpreting this information. Hence, we expect stocks with higher return volatility to have higher levels of information asymmetry.

2.2.2.4 Analysts following and forecast dispersion

The number of analysts following a firm is a commonly used measure of the amount of publicly available information about the firm. It has also been proposed as a measure of private information. For instance, Brennan and Subrahmanyam (1995) found that an increase in the number of analysts following a stock is associated with a drop in information asymmetry. Thus, increased competition among informed traders (measured by the number of analysts) lowers the adverse selection problem faced by market-makers. Brennan, Jegadeesh, and Swaminathan (1993) found that the stock prices of firms with more analysts react more rapidly to news than do the prices of firms with fewer analysts.

The cross-sectional standard deviation of analyst forecasts is used as a measure of uncertainty regarding firm fundamentals, and hence serves as a natural measure of Σ_0. (Coller and Yohn, 1997). Since market-makers facing higher uncertainty update prices to a greater extent in response to order-flow, we expect to find a positive relation between information asymmetry and the variability in analyst forecasts.

2.2.2.5 Institutional ownership

Institutional shareholders are considered to be investors more likely to possess private information. Firms may voluntarily reveal private information to a small group of investors (Bhattacharya and Chiesa, 1995; Yosha, 1995). Moreover, institutional share holders are in a better position to monitor a firm, and thus, have an informational advantage over other investors.

2.2.2.6 Firm characteristics

Market-to-book (MB) is a measure of a firm's growth options. Therefore, high MB firms are expected to display higher information asymmetry (Smith and Watts, 1992). Firm size is an inverse proxy for the existence of private information. For instance, Zeghal (1983) showed that the information content of financial statements is lower for large firms because most of it has already been impounded by the market, while Seyhun (1986) found that the profits to insider trading are inversely related to firm size. Information asymmetry is expected to be lower for larger firms.

2.3 Empirical Methods and Construction of Variables

2.3.1 Measuring information asymmetry

Several papers have developed measures of information asymmetry.[4] Underlying each is the intuition that trading on private information should have a permanent effect on prices. We compute measures of information asymmetry, as proposed by Glosten and Harris (1988), Hasbrouck (1991), George, Kaul and Nimalendaran (1991), and Lin, Sanger and Booth (1995). Table 2.1 provides the time-series average of the weekly cross sectional correlations among these four estimates of information asymmetry. The correlations are high, and hence, for the sake of brevity, we only report the results using the Glosten and Harris (1988) measure. Our conclusions are robust to the choice of measure.

In the Glosten and Harris (1988) model, price change is decomposed into a permanent component due to adverse selection (information asymmetry) and a transitory component due to inventory and order-processing costs. Specifically, the change in transaction price is related to transaction volume and to a buy/sell indicator in the following reduced form specification:

$$\Delta P_t = c_0 \Delta I_t + c_1 \Delta I_t V_t + z_0 I_t + z_1 I_t V_t + \varepsilon_t \qquad (2.1).$$

---

[4] See, among many others, Glosten (1987); Glosten and Harris (1988); Hasbrouck (1988); Stoll (1989); George, Kaul, and Nimalendran (1991); Hasbrouck (1991); Madhavan and Smidt (1991); Huang and Stoll (1997); and Madhavan, Richardson, and Roomans (1997).

Here, I_t is a trade indicator that equals 1 if the t^{th} transaction is buyer-initiated and -1 if it is seller-initiated; P_t is the transaction price for the t^{th} trade; V_t is the volume traded; and ε_t captures public news. In this model, the adverse-selection (information asymmetry) component is $2\left(z_0 + z_1 V_t\right)$, and the inventory-holding and order-processing components are together captured by $2\left(c_0 + c_1 V_t\right)$.[5]

We collect trade and quote data for each IPO and estimate model (1) in every week or month, depending on the frequency of the analysis. Once we have the regression coefficients for the stock, we can calculate the components of the spread. Since the two components depend on volume, we use the average transaction size for a stock to obtain estimates of the components. Additionally, to compare estimates across stocks, we express the adverse selection component as a percentage of the spread:

$$IA = \frac{2\left(z_0 + z_1 \overline{V}\right)}{2\left(c_0 + c_1 \overline{V}\right) + 2\left(z_0 + z_1 \overline{V}\right)} \tag{2.2}$$

where $\overline{V}$ is average transaction volume. We follow the Lee-Ready (1991) procedure for classifying trades. According to this algorithm, a trade is classified as buyer- (seller-) initiated if the transaction price is closer to the ask (bid) price of the prevailing quote. The quote must be at least five seconds old. If the trade occurs at the midpoint of the quote, the "tick test" is employed. In this case, a trade is classified as a buy if the previous price change is positive, and is classified as a sell if it is negative. Since trade direction is inferred from the data, some assignment error inevitably occurs. Nevertheless, as shown by Lee and Radhakrishna (2000) and Odders-White (2000), the algorithm is largely accurate.

---

[5] The model is expressed in terms of transaction price changes. The implied bid and ask prices are obtained by substituting I_t = -1 and I_t = +1 into (1).

2.3.2 *Independent variables*

(i) Order-flow variability. Using the Lee-Ready algorithm, we compile 15-minute order-flow for each stock and then calculate its standard deviation for each week or month. Order-flow can be measured in terms of net number of trades (number of buys less number of sells in each 15-minute period), net volume (buy volume, less sell volume) or net value (value of buys, less value of sells). Accordingly, we have three possible measures of order-flow variability. We focus on the standard deviation of net number of trades; using the standard deviation of net volume and net value yields similar inferences. For the sake of brevity, we refer to the variability of 15-minute order-flow as order-flow variability, or σ_{OF}. The TAQ database is the source of the order-flow data.

(ii) Proportion of small and medium trades. We use Barclay and Warner (1993) to define trades of less than 500 shares as small; trades of between 500 and 9,900 shares as medium; and trades in excess of 10,000 shares as large. For each stock, we calculate the fraction of weekly or monthly volume or total number of trades occurring in trades of less than 500 shares (small), trades of between 500 and 9,900 shares (medium), and trades in excess of 10,000 shares (large).

(iii) Return volatility. To measure weekly return volatility, we use the Parkinson (1980) range-based measure of volatility. The return volatility of firm i, in week t, is $V_{i,t} = \ln\left(\frac{P_{i,high,t}}{P_{i,low,t}}\right)$. Here $P_{i,high,t}$ and $P_{i,low,t}$ are the maximum and minimum prices for stock

i, in week t. When measuring monthly volatility, we estimate idiosyncratic volatility following Ang, Hodrick, Xing, and Zhang (2004). For each month, we run the following regression for firms with more than 17 daily return observations within that month:

$$r_{i,t,d} = \alpha_{i,t} + \beta_{i,m} \times r_{m,d} + \beta_{i,SMB} \times SMB_{t,d} + \beta_{i,HML} \times HML_{t,d} + \varepsilon_{i,t,d} \qquad (2.3)$$

where, for day d, in month t, $r_{i,t,d}$ is stock i's excess return; $r_{m,d}$ is the excess return on the market portfolio; and $SMB_{t,d}$ and $HML_{t,d}$ are the Fama-French (1993) size and book-to-market factor returns. $\varepsilon_{i,t,d}$ is the residual, with respect to the Fama-French (1993) three-factor model. We use the standard deviation of the daily residuals in month t to measure the idiosyncratic volatility for firm i, in month t.

(iv) Market information asymmetry (IA_{mkt}). We calculate the average information asymmetry across all stocks, but separately for the NYSE and NASDAQ. Before using the index in any regression, we adjust it to remove the value of information asymmetry for the firm under consideration. We use this variable to test if issuers time the market to take advantage of periods of low overall information asymmetry in the market.

(v) Firm size and market-to-book(MB). Size is measured using price and number of shares outstanding at the end of each week (month). *MB* is computed using the Compustat quarterly file.

(vi) Analyst and institutional ownership data. Analysts following and forecast data are obtained from the I/B/E/S tapes. Institutional ownership data are obtained from the

Spectrum database. Spectrum records the SEC mandated, 13-F filings of large institutional investors, that provides quarterly snapshots of institutional holdings.

2.3.3 Model specifications

To understand the cross-sectional influences on information asymmetry, we estimate the following cross-sectional regression model in each event week, starting at week 2, and once again, separately for NASDAQ and the NYSE:[6]

$$\begin{aligned} IA_{i,t} = {} & \alpha_t + \beta_{1,t} \times \sigma_{OF,i,t-1} + \beta_{2,t} \times PT_{small,i,t-1} + \beta_{3,t} \times PT_{medium,i,t-1} + \beta_{4,t} \times \ln\left(Price_{i,t-1}\right) + \\ & \beta_{5,t} \times IA_{mkt,t-1} + \beta_{6,t} \times \ln\left(Vol_{i,t-1}\right) + \beta_{7,t} \times \ln(Size_{i,t-1}) + \beta_{8,t} \times Volatility_{i,t-1} + \varepsilon_{i,t} \end{aligned} \tag{2.4}$$

where:

- $IA_{i,t}$ is the GH measure of information asymmetry for firm i in week t.
- $Volatility_{i,t-1}$ is the previous week's return volatility based on Parkinson (1980). We expect the coefficient to be positive, if stocks with higher volatility are characterized by greater information arrival and informed trading.
- PT_{small} and PT_{medium} are the proportion of small trades (trades of less than 500 shares) and medium trades (500-9,900 share trades) in the previous week. Assuming that small trades come mainly from uninformed investors and medium trades come from informed investors, we expect β_2 to be negative and β_3 to be positive.
- IA_{mkt} is the equally weighted average market information asymmetry, adjusted for the dependent firm's information asymmetry.

---

[6] The independent variables for week 2 might be for a partial week 1.

- $\sigma_{i,OF,t-1}$ is the previous week's order-flow variability. The coefficient is expected to be negative, so that $\sigma_{i,OF,t-1}$ is a measure of the extent of uninformed trading.

- Size is the natural logarithm of the market capitalization of the IPO firm at the end of the previous week. The coefficient is expected to be negative, as information asymmetry should be more pronounced for smaller firms.

- $\ln\left(Price_{i,t-1}\right)$ is the average transaction price in the previous week. The GH measure of the adverse selection component of the spread is computed as a fraction of the spread, and could depend on the size of the spread. Since the spread is, in turn, a function of the stock price, price levels must be controlled before comparing information asymmetry across stocks.

- $\ln\left(Vol_{i,t-1}\right)$ is trading volume in the previous week. Volume is often used as a proxy for liquidity, and thus might have additional explanatory power for the level of information asymmetry.

Specification (4) focuses on the association between information asymmetry and the firms' trading environment. To examine the effects of firm characteristics in addition to these trading characteristics, we estimate the following cross-sectional regression model in each event month t:

$$\begin{aligned} IA_{i,t} = {} & \beta_{0,t} + \beta_{1,t} \times \ln\left(Inst_{i,t-1}\right) + \beta_{2,t} \times \ln\left(Size_{i,t-1}\right) + \beta_{3,t} \times \sigma_{OF,i,t-1} + \beta_{4,t} \times PT_{small,i,t-1} + \\ & \beta_{5,t} \times MB_{i,t-1} + \beta_{6,t} \times \ln\left(Analyst_{i,t-1}\right) + \beta_{7,t} \times \sigma_{res,i,t-1} + \beta_{8,t} \times \sigma_{forecast,i,t-1} + \varepsilon_{i,t} \end{aligned} \tag{2.5}$$

where the additional independent variables are as follows:

- $Inst_{i,t-1}$ is the number of institutional shareholders.
- $Analyst_{i,t-1}$ is the number of institutional analysts following the firm.
- $\sigma_{forecast,i,t-1}$ is the dispersion in their forecasts.
- $\sigma_{res,i,t-1}$ is the idiosyncratic return volatility.

Since these variables are all measured at lower frequencies, we use monthly analogs to the weekly variables and extend our horizon of study from two years to approximately four years. Specifically, we drop the first calendar month after the IPO and calculate information asymmetry in eight event months – 2, 8, 14, 20, 26, 32, 38 and 44 – relative to the IPO date. The six-month separation between event months is dictated by the availability of Compustat, I/B/E/S and the ownership data. We compute the standard deviation of 15-minute order-flow and the proportions of small and medium traders in the month before the month in which information asymmetry is calculated. We also compile the analyst and institutional investor data as of the end of the month before the event month.

2.4 Sample Selection and Sample Characteristics

2.4.1 Sample selection and data

For inclusion in our sample, we require a firm to have four years of trading history post-IPO and an offer price greater than $10. The sample consists of 2,132 firms with an IPO on NYSE/AMEX (henceforth, for brevity, NYSE) or NASDAQ, between January 1993 and December 1998.[7] We exclude 130 financial sector firms, leaving a sample of 2,002 firms.[8] The list of firms and the dates of their IPOs are obtained from the SDC Platinum database.

---

[7] The choice of sample period (January 1993 to December 1998) is determined by the availability of TAQ data on the University of Alberta Finance Server (Lorax) (1993-2002).

The Center for Research in Security Prices (CRSP) database is used to obtain daily stock returns and daily volume data, as well as the exchange on which the stock trades, and its four-digit Standard Industrial Classification (SIC) code. Non-availability of 52 firm records in the CRSP database reduces the sample to 1950 (1661 NASDAQ firms and 289 NYSE firms). Intra-day data for the firms was obtained from Trade and Quote (TAQ) CDs. Following Chordia, Roll, and Subrahmanyam (2001), several filters were employed to ensure the validity of the TAQ data. The first trade of each day is dropped from the analysis, since opening trades usually occur through a call auction.

We calculate the measures of information asymmetry in every event week during the first two years following the IPO. In addition, for the purposes of our cross-sectional analysis, we recalculate the information asymmetry measures in eight event months, (months 2, 8, 14, 20, 26, 32, 38 and 44), relative to the IPO date.

Our cross-sectional tests relate information asymmetry to several firm-specific variables, as described in Section 3.3. Having excluded firms with missing data, we use a sample of 1,100 firms in the cross-sectional analysis. Dropping the analyst variables increases the sample to about 1,700 in each event month. To test the sensitivity of our results to the requirement that analyst data be available, we re-estimate our regression specifications using the smaller sample. The coefficients on the remaining variables behave as in the full sample of 1,700 firms.

---

[8] Financial sector firms are both highly regulated and highly levered, which make it hard to interpret the adverse selection component estimates.

We follow Hand and Skantz (1999) in extracting, from the SDC database, a sample of 171 equity carve-outs issued between January 1993 and December 1998 (77 of these trade on the NYSE, while 94 trade on NASDAQ).[9] As with the IPO sample, we compute weekly adverse selection costs and order-flow variability for the 171 carve-outs for two years post-IPO.

2.4.2 *Sample characteristics*

Summary statistics for the sample of IPO firms are provided in Tables 2.1(A) and 2.2(B). As shown in Table 2.2(A), the average firm in the sample has a market capitalization of $260 million (based on the day one closing price). The average size of IPO firms on NASDAQ ($183 million) is roughly one-quarter the size of the average IPO on the NYSE ($700 million). This difference in size is statistically significant. The average NYSE firm has 3,612 employees, which is roughly six times the number of employees in the average NASDAQ firm (597). The average firm listing its IPO on the NYSE is older (2,736 days) than the corresponding firm on NASDAQ (2,048 days), though the difference is not statistically significant.[10]

Offered shares and offered proceeds average $13.81 million and $250.4 million on the NYSE and $3.7 million and $48.98 on NASDAQ. These numbers are in concurrence with the earlier findings of Corwin, Harris, and Lipson (2004) and Ellis, Michaely, and O'Hara (2002). The average IPO float accounts for a little over 30% of the outstanding shares in each market. While 15% of the IPO float on the NYSE comes from pre-IPO shareholders (i.e., the company sells the remaining 85%), the corresponding figure for NASDAQ IPOs is 8%. The average IPO on the NYSE has a filing range of about 15% while that on NASDAQ has a range of about 18%.

---

[9] An initial screen required SDC's spinoff code be set to "Yes" and the units code be set to "No". SDC's spinoff code is #438. It is set to "Yes" when the issue is deemed by SDC to occur "when a company decides to distribute shares representing ownership in a division or subsidiary of the company that will now trade separately from its former parent." Transactions labelled spinoffs in the Worldwide New Issues Database are therefore public offerings, not true spinoffs (*pro rata* distributions of subsidiary stock to parent shareholders). The SDC's units code is #940. It is set to "Yes" when the offering is for units. Since a unit represents a combination of securities such as common stock, debt, preferred stock, and warrants, unit public offerings may have very different economic characteristics and be issued by firms for different reasons than all-equity carve-outs. We therefore excluded all subsidiary IPOs that were units.

[10] Firm age data are obtained using the incorporation date value in SDC. This variable is omitted from the reported cross-sectional regression models because it is not significant in any month.

Average order-flow variability on NASDAQ (9.48 trades) is higher than on the NYSE (5.3 trades). This could be due to the large volume of inter-dealer transactions on NASDAQ. To the extent that order-flow variability is a measure of the extent of uninformed trading, this result implies a greater level of uninformed trading for NASDAQ IPOs relative to the NYSE IPOs. The result is important, and as we show in Section 3.6, it helps in explaining the patterns in information asymmetry.

The day one average bid ask spread for NASDAQ IPOs is 31 cents while the corresponding spread for the NYSE IPOs is 15 cents. This difference, significant for both buyer- as well as seller-initiated trades, is consistent with the findings of Falconieri, Weaver, and Murphy (2004). As with Corwin et al. (2004) and Falconieri et al. (2004), we observe substantial differences in order-flow between NASDAQ and the NYSE IPOs. The average day one buyer-initiated and seller-initiated trades are significantly larger on the NYSE than on NASDAQ.

2.5 Information Asymmetry in the IPO Aftermarket

The variation through time in weekly information asymmetry, as measured by the Glosten and Harris (1988) decomposition (Equation (2.2)), is summarized in Table 2.4

and Figure 2.1(A). We track each firm for two years after its IPO and calculate information asymmetry on a weekly basis. We drop the week (often a partial week) in which the IPO takes place. The results are presented separately for the NYSE and NASDAQ stocks. For comparative purposes, Figure 2.1(B) presents the evolution of the mean IPO spread as a function of time.

Several results are of interest. First, information asymmetry is lowest in the second week of trading, with the mean being 0.18 (i.e., 18% of the spread) for the NYSE stocks and 0.08 for NASDAQ stocks. It then increases steadily for the next 8 to 12 weeks for both the NYSE and NASDAQ stocks, before settling at a level of approximately 0.32 for the NYSE stocks and 0.12 for NASDAQ stocks. This pattern is reinforced by that in the raw spread (a coarser measure of adverse selection). As shown in Figure 2.1(B), the average quoted spread increases for 10 weeks for the NYSE stocks and for 13 weeks for NASDAQ stocks. The mean spread reaches a maximum of $0.21 for the NYSE stocks and declines to $0.18 after 100 weeks, whereas, for NASDAQ stocks, it declines from a maximum of $0.42 to approximately $0.29 after 100 weeks. The fact that the initial increase is similar for the GH measure and the raw spread indicates that the pattern in adverse selection is not sensitive to our use of a specific measure (the GH measure) of information asymmetry.[11]

At first glance, the low information asymmetry in the second week is counterintuitive. With virtually no public history, information asymmetry about the value of the firm should be high. Nevertheless, our measure of information asymmetry reflects the information asymmetry between investors, which will be low if all (or most) investors are uninformed. It is plausible that investors in the IPO aftermarket are largely uninformed.

---

[11] We find similar patterns in the evolution of post-IPO adverse selection using the measures proposed by George, Kaul and Nimalendaran (1991), Lin, Sanger and Booth (1995), and Hasbrouck (1991).

We examine the changes in information asymmetry around the time that various post-IPO regulations are lifted. The first of these events is the end of the quiet period in week four (25 calendar days, post-offering), which sees the beginning of analyst and financial press coverage. We see an increase in information asymmetry between week two and week four. Table 2.4 shows that for both the NYSE and NASDAQ IPOs, about 50% of the total increase in information asymmetry takes place between week two and week four. Nevertheless, only 10% of the increase in information asymmetry for the NYSE IPOs and about 20% of the increase for NASDAQ IPOs takes place between weeks three and four. Thus, the end of the quiet period explains a small fraction of the increase in information asymmetry.

To discourage investors from selling their IPO shares immediately after the offering, underwriters impose penalty bids for "flipping" the stock (Aggarwal, 2000). A large proportion of the initial share allocation is to sophisticated (thus, potentially informed) institutional investors (Brennan and Franks, 1997). Consequently, a significant increase could occur in information asymmetry around the time that penalty bids are lifted, typically, about 30 days (six weeks) after the offering. Table 2.4 shows that about 23% of the total increase in information asymmetry for the NYSE IPOs (15% for NASDAQ IPOs) takes place in week six. Thus, penalty bids can also explain a small fraction of the initial increase in information asymmetry.

Rule 144 prevents officers and directors of IPO firms from selling their shares within 90 days of the IPO, and the subsequent entry of officers and directors could increase information asymmetry.[12] Although we do not see an increase between week 12 and week 13, information asymmetry plateaus after week 13. Additionally, most IPOs have lockup provisions that restrict managers, large investors, and existing investors in the issuing firms from selling their shares, typically, for 180 days (or 26 weeks). Field and Hanka (2001) report that a permanent increase occurs in volume following the lockup expiration date, and they attribute this to selling, on the part of managers and other restricted investors. We expect trades by managers and large investors to increase information asymmetry. Nevertheless, our analysis shows that information asymmetry in weeks 26 and 27 is similar to its value in week 12. If the increase in volume associated with the end of the lockup period is driven by a combination of informed trades and uninformed trades (e.g., managers selling shares for portfolio diversification purposes), then information asymmetry will not worsen.

While our results suggest that the lifting of regulations affects the level of information asymmetry (evidence in support of H_4), questions about the slope of its increase remain unanswered. We expect to see sharp jumps in week 4 (the end of the quiet period), week 6 (the termination of penalty bids), week 13 (the commencement of sales under rule 144), and week 26 (the end of the lock-up period), with information asymmetry being relatively flat in other weeks. Figure 2.1 and Table 2.4 show a fairly smooth increase in information asymmetry through weeks 8 to 12.

---

[12] Rule 701 prevents non-affiliates from selling their pre-IPO allocation within 90 days of the offering. The relaxation of this rule is unlikely to have an impact on information asymmetry, since sales from non-affiliates are likely to occur for liquidity or diversification reasons (Cao, Field, and Hanka, 2004).

We examine explanations for the absence of the expected step patterns in information asymmetry. Aggarwal et al. (2002) show that uninformed traders are attracted to IPOs due to underpricing and media coverage, and are heavy traders of the IPO shares for a few weeks. The dominance of uninformed traders in the immediate IPO aftermarket would explain the initially low levels of information asymmetry. If these traders eventually exit the market, the fraction of informed traders will increase, as will information asymmetry.

To address this possibility, we examine the pattern in order-flow variability (σ_{OF}) that, following Kyle (1985), is a natural measure of the extent of uninformed trading. As shown in Figure 2.2, the first two weeks in the post-IPO life of the firm are characterized by unusually high σ_{OF}. This figure also shows a steady decline in σ_{OF} between week two and week eight. Thus, uninformed traders appear to be active in the early IPO aftermarket, and the extent of uninformed trading declines as the firm becomes more seasoned.

Figure 2.3 plots the mean proportion of medium (500-9,900 share) trades in each event week. This is a proxy for informed trading. A steady increase is seen in the mean proportion of weekly trades and volume, occurring in medium trades over the initial weeks of trading (evidence in support of H_3). For instance, over the first 12 weeks of trading, the mean fraction of medium trades increases from 56% to 65% for the NYSE IPOs, and from 71% to 80% for NASDAQ IPOs. This is consistent with an increase in informed trading in the early phase of post-IPO trading.

The patterns in informed and uninformed trading fit those of information asymmetry. Specifically, uninformed trading is relatively heavy immediately after the IPO, which explains the initial low levels of information asymmetry. As the extent of uninformed trading declines, informed investors are unable to hide their trades as effectively and information asymmetry increases. This effect is reinforced by the relaxation of post-IPO regulations, which serve to increase the extent of informed trading.

To formally confirm the relation between information asymmetry and order-flow variability, and the proportion of medium trades, we estimate a time-series variant of Equation (2.4). The model is estimated for each firm in the sample. To avoid issues of simultaneity and endogeneity, we use lagged independent variables. Table 2.5 presents the cross-sectional average of the coefficients. Taking into consideration the non-linear pattern in the evolution of information asymmetry (Figure 2.1), we divide the post-IPO two-year analysis period into two sub-periods. The first period consists of week 2 through week 52 and the second period spans week 53 to week 104. The results for the first sub-period for the NYSE and NASDAQ are presented in panels A and B, respectively, while panels C and D present the results for the second sub-period.

The results are consistent across sub-periods one (weeks 2 through 52) and two (weeks 53 through 104). Information asymmetry is consistently negatively related to both σ_{OF} and the proportion of small trades (evidence in support of H_2). Thus, periods of market-wide high variability in order-flow and a larger proportion of small trades are marked by lower information asymmetry. We find a consistent negative relation between trading volume and information asymmetry. Since trading volume is a measure of liquidity, and information asymmetry captures one component of transaction costs, the observed negative relation is reiterating the inverse relation between liquidity and transaction costs. The coefficient for idiosyncratic return volatility is positive, suggesting that a high level of firm-specific information arrival leads to higher information risk for the market-maker.

We find a positive (marginally) significant relation between market-wide information asymmetry and firm information asymmetry. This result points to the existence of a systematic component in the adverse selection cost of trading.

We explore the importance of prior information ($H_{1,1}$, $H_{1,2}$ and $H_{1,3}$) by comparing IPO aftermarket information asymmetry with information asymmetry in the carve-out aftermarket. Table 2.6 presents the results. In the absence of any controls, the average level of information asymmetry for the carve-outs is higher than that for the new IPOs. The higher level of information asymmetry for carve-outs suggests that investors with superior processing ability are able to exploit the greater information available for carve-outs. Controlling for price, volume, and idiosyncratic return volatility, the post-issue information asymmetry for carve-outs and IPOs is identical. Similar to the patterns observed in the IPO aftermarket, the adverse selection cost of trading in the carve-out aftermarket is lowest immediately post-issuance and increases for 8 to 12 weeks before stabilizing. The similar pattern in information asymmetry for IPOs and carve-outs suggests that this pattern results from variation, since it is largely a result of the extent of informed vs. uninformed trading.

A comparison of information asymmetry for the NYSE and NASDAQ IPOs reveals that the level of adverse selection is larger for the NYSE IPOs than for NASDAQ IPOs (Figure 2.1(A)). The NYSE-NASDAQ differential is consistent with earlier findings by Lee (1993), Seppi (1990), Affleck-Graves, Hegde and Miller (1994) and Lin, Sanger and Booth (1995). As a possible explanation, a large fraction of the volume in NASDAQ stocks comprises inter-dealer transactions (Ellis et al., 2002), for which adverse selection is low. Another possible reason may that NASDAQ dealers are better informed about the firms in whose shares they make a market. For instance, the lead underwriter for NASDAQ IPOs is almost always the main market-maker for the stock, and close relations with the firm might lower the adverse selection problem for NASDAQ IPOs. This issue, while interesting, is not central to our purposes. We are concerned with the evolution of information asymmetry in the IPO aftermarket, which is identical for the NYSE and NASDAQ IPOs.

2.6 The Determinants of Information Asymmetry

Table 2.3 shows that the cross-sectional standard deviation of information asymmetry (computed by event week) is between one-quarter and one-third of the mean level of information asymmetry for the NYSE IPOs, and between 65% and 70% of mean information asymmetry for NASDAQ IPOs. Thus, considerable variation exists in information asymmetry around the mean in every event week, for both the NYSE and NASDAQ IPOs. In this section, we study the extent to which trading and firm characteristics can explain this cross-sectional variation in information asymmetry.

2.6.1 Trading characteristics and information asymmetry

The model specified in Equation (2.4) is estimated separately for NASDAQ and the NYSE IPOs, and the results are presented in Table 2.7. Due to a high correlation between the PT_{small} and PT_{medium}, we estimate the model twice to separately consider the effects of PT_{small} and PT_{medium}.

Order-flow variability, PT_{small}, PT_{medium}, and return volatility stand out as the set of significant explanatory variables. The coefficient on order-flow variability is negative. Thus, stocks with greater order-flow variability have lower information asymmetry. This is consistent with microstructure models (e.g., Kyle, 1985) which predict that higher levels of order-flow variability allow informed traders to hide and result in lower levels of information asymmetry. The coefficient on PT_{small} is negative, while the coefficient of PT_{medium} is positive, consistent with greater levels of uninformed trading (a higher proportion of small trades) or lower levels of informed trading (fewer medium trades), reducing the level of information asymmetry.

The coefficient on stock return volatility is consistently positive. Since return volatility is a proxy for information arrival, this coefficient suggests that firms with greater information arrival are characterized by more informed trading and hence higher information asymmetry. For the NYSE IPOs, information asymmetry is negatively related to market capitalization in each event week, except week two. Thus, smaller IPOs have higher levels of asymmetry, which is consistent with the conventional wisdom that small stocks face a greater likelihood of informed trading. The coefficient is also negative for NASDAQ IPOs, but it is not always significant. While the coefficient on trading volume is negative for both the NYSE and NASDAQ IPOs – consistent with the notion that more liquid assets have lower levels of adverse selection – it is not consistently significant in either sample.

We examine the patterns in the weekly coefficients across the two years post-IPO. No clear trends are apparent in any of the coefficients, though we find that the coefficient on σ_{OF} increases around the expiration of the quiet period and the lifting of penalty bids. This suggests that these periods are unusual, plausibly due to a change in the amount of informed trading during these weeks.

2.6.2 *Firm characteristics and information asymmetry*

Equation (2.5) presents an augmented specification with firm level variables including the number of analysts, the dispersion in their forecasts, the number of institutional shareholders, and the market to book ratio. Table 2.8 presents the results.

Consistent with the view that analysts mitigate information asymmetry, we find that the coefficient for the number of analysts following the IPO is negative (though it is not significant). The coefficient for earnings forecast dispersion is not significant. We find that the coefficient for the number of institutional investors is significantly above zero in month two; however, for all other months, the coefficients are negative and insignificantly different from zero. The positive coefficient in month two could indicate that the market is more wary of institutional shareholders in the early stages of post-IPO trading. The coefficient for MB, while positive in most months, is never statistically significant.

Consistent with the results in Table 2.7, the residual return volatility coefficient is positive and significant. This suggests that greater firm-specific information arrival increases chances of informed tradeing, thereby raising information asymmetry. Also, as before, the coefficient for order-flow variability is significantly below zero in every event month. Thus, the role of uninformed traders in lowering information asymmetry is robust to the inclusion of firm-specific variables. In a separate specification, we include PT_{small} instead of σ_{OF} (the average cross-sectional correlation between monthly σ_{OF} and PT_{small} is 0.82), and find a negative coefficient. Thus, as before, information asymmetry declines for stocks with higher levels of uninformed trading.

2.7 Conclusion

This paper explores the evolution of information asymmetry in the IPO aftermarket. We use the adverse selection component of the spread as a proxy for information asymmetry. Information asymmetry is of interest to firms since it has been shown to affect their cost of capital, and could have implications for the success or failure of a new issue. Moreover, since the adverse selection component of the spread is a measure of the permanent price impact of trades, our study yields insights into the price formation process in the IPO aftermarket.

Our time-series analysis provides the surprising result that information asymmetry is at its lowest level immediately post-IPO, and increases for about 8 to 12 weeks before stabilizing. It does not appear as if this pattern in information asymmetry is directly associated with post-IPO regulations. Rather, our analysis suggests that order-flow variability (a measure of uninformed trading) and the proportion of small vs. medium trades (a measure of the relative presence of uninformed vs. informed investors) play important roles in the post-IPO pattern of information asymmetry.

Overall, the initial increase in information asymmetry can be explained in terms of the following effects: first, the lack of public information in the early aftermarket prevents investors with superior processing ability from generating any private advantages. With the passage of time, as the firm matures and the aftermarket regulations are lifted, more public and private information is generated. Consistent with the models of Kim and Verrecchia (1994) and Kandel and Pearson (1995), this increased information set provides traders with more opportunities to generate an informational advantage. Second, as a trading history develops, investors learn the properties of order-flow. This causes informed investors to initially trade less aggressively and increase their trading intensity over time (as their confidence in order-flow properties increases). Finally, uninformed trading activity is initially heavy, as uninformed traders chase recent IPOs. This provides informed traders with camouflage. Over time, uninformed trading decreases and stabilizes and this reduces the ability of informed traders to hide.

The cross-sectional analysis shows that order-flow variability and the fraction of medium trades, together with return variability, are the important determinants of inter-firm variations in information asymmetry. After controlling for these variables, we find that the role of analysts and institutional shareholders in the evolution of the aftermarket information asymmetry is weak. Interpreted in the context of microstructure models such as Kyle (1985), our results suggest that direct measures of the intensity of uninformed trading (order-flow variability or the proportion of small trades) and of informed trading (the proportion of medium trades or idiosyncratic return variance) are better at explaining cross-sectional variations in information asymmetry than are firm characteristics.

References

Affleck-Graves, John, Shantaram P. Hegde and Robert E. Miller, 1994. Trading mechanisms and the components of the bid-ask spread, *Journal of Finance* 49, 1471-1488.

Aggarwal, R., 2000. Stabilization Activities by Underwriters after Initial Public Offerings, *Journal of Finance*, 55, 1075–1103.

Aggarwal, R.K., L. Krigman, and K.L. Womack, 2002. Strategic IPO underpricing, information momentum, and lockup expiration selling, *Journal of Financial Economics* 66, 105– 137.

Aggarwal, Reena, Ricardo Leal and Leonardo Hernandez 1993. The Aftermarket Performance of Initial Public Offerings in Latin America, *Financial Management* 22, 42-53.

Aggarwal, R. and P. Rivoli, 1990. Fads in the Initial Public Offering Market? *Financial Management* 19, 45–57.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2004. The cross-section of volatility and expected returns, *Journal of Finance*, 61, 259-299

Atiase, R.K., 1985. Pre-disclosure information, firm capitalization, and security price behavior around earnings announcements. *Journal of Accounting Research*, 23(1): 21-36.

Barclay, M. J. and J. B. Warner, 1993. Stealth trading and volatility. Which trades move prices? *Journal of Financial Economics* 34, 281-305.

Bhattacharya, S., and Chiesa, G., 1995. Proprietary Information, Financial intermediation, and research incentives. *Journal of Financial Intermediation* 4, 328-357.

Brav, A. and P. Gompers, 2003. The Role of Lockups in Initial Public Offerings, *The Review of Financial studies* 16, 1-29.

Brennan, Michael, and Julian Franks, 1997. Underpricing, ownership and control in initial public offerings of equity securities in the UK, *Journal of Financial Economics* 45, 391–413.

Brennan, M.J., Jegadeesh, and B. Swaminathan, 1993. Investment Analysis and the adjustment of Stock Prices to Common Information, *Review of Financial Studies* 6, 799-824.

Brennan, M. J. and A. Subrahmanyam, 1995. Investment analysis and price formation in securities markets, *Journal of Financial Economics* 38, 361-381.

Brennan, M.J. and A. Subrahmanyam, 1996. Market Microstructure and Asset Pricing: On the Compensation for Illiquidity in Stock Returns. *Journal of Financial Economics* 41, 441-464.

Cao, Charles, Laura C. Field and Gordon Hanka, 2004. Does insider trading impair market liquidity? Evidence from IPO lockup expiration, *Journal of Financial and Quantitative Analysis*, 39, 25-46.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2001. Market liquidity and trading activity, *Journal of Finance* 56, 501-30.

Coller, M., and Yohn, T.L., 1997. Management forecasts and information asymmetry: An examination of bid-ask spreads. *Journal of Accounting Research*, 35(2): 181.

Corwin, Shane A., Jeffrey H. Harris, and Marc L. Lipson, 2004. The Development of Secondary Market Liquidity for NYSE-Listed IPOs, *Journal of Finance*, 59, 2339-2373.

Easley, D., and O'Hara, M., 1987. Price, trade size and information in securities markets. Journal of Financial Economics 19(1): 69-90.

Easley, D., S. Hvidkjaer and O'Hara M., 2002. Is information risk a determinant of asset returns? Journal of Finance 57, 5, 2185-2221.

Ellis, Katrina, Roni Michaely and Maureen O'Hara, 2002. The making of a dealer market: From entry to equilibrium in the trading of Nasdaq stocks, *The Journal of Finance* 57, 2289-2316.

Falconieri, S., D. Weaver, and A. Murphy, 2004. From IPO to the first trade: Is underpricing related to the trading mechanism, Working paper, Tilburg University.

Fama, Eugene F. and Kenneth R. French, 1993. Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33 (1), 3-56.

Foster F. D. and S. Viswanathan, 1993. Variations in trading volume, return volatility, and trading costs: Evidence on recent price formation models, *Journal of Finance* 48, 187-211.

Field , L.C. and G. Hanka, 2001. The Expiration of IPO Share Lockups, *Journal of Finance*, 56, 471-500.

Froot, K. A., 1989. Consistent covariance matrix estimation with cross-sectional dependence and heteroskedasticity in financial data, *Journal of Financial and Quantitative Analysis* 24, 333–355.

George, T.J., G. Kaul, and M. Nimalendran, 1991. Estimation of the bid-ask spread and its components: A new approach, *Review of Financial Studies* 4, 623-656.

Glosten, Lawrence R., 1987. Components of the bid-ask spread and the statistical properties of transaction prices, *Journal of Finance* 42, 1293-1307.

Glosten, L.R. and L. E. Harris, 1988. Estimating the components of the bid/ask spread, *Journal of Financial Economics*, 21, 123-142.

Hand, J., and T. Skantz, 1999. Tax planning in initial public offerings: The case of equity carve-outs, Working Paper, University of North Carolina at Chapel Hill and Florida Atlantic University.

Hasbrouck, J., 1988. Trades, Quotes, Inventories, and Information, *Journal of Financial Economics* 22, 229-252.

Hasbrouck, J., 1991. Measuring the information content of stock trades, *Journal of Finance* 46, 129-207.

Huang, R.D. and H.R. Stoll, 1997. The components of the bid-ask spread: A general approach, *The Review of Financial Studies* 10, 995-1034.

Huson, Mark R., and MacKinnon, G., 2003. Corporate spinoffs and information asymmetry between investors. *Journal of Corporate Finance*, 9(4): 481-503.

Kandel, E., and N.D. Pearson, 1995. Differential Interpretation of Public Signals and Trade in Speculative Markets, *Journal of Political Economy*, 103, pp.831-872.

Kim, O., and Verrecchia, R.E., 1994. Market liquidity and volume around earnings announcements. *Journal of Accounting and Economics*, 17(1-2): 41-67.

Kyle, A., 1985. Continuous auctions and insider trading, *Econometrica* 53, 1315-1336.

Lang, Mark, 1991. Time-varying stock price response to earnings induced by uncertainty about the time-series process of earnings. *Journal of Accounting Research* 29, 229-257.

Lee, C.M.C., 1993. Market Integration and Price Execution for NYSE Securities, *Journal of Finance*, 48, 1009-1038.

Lee, C., and M. Ready, 1991. Inferring trade direction from intraday data, *The Journal of Finance* 46, 733-746.

Lin, J. -C., G. C. Sanger, and G. G. Booth, 1995. Trade size and components of the bid-ask spread, *The Review of Financial Studies* 8, 1153-1183.

Madhavan, A., M. Richardson, and M. Roomans, 1997. Why do security prices change? A transaction-level analysis of NYSE stocks, *The Review of Financial Studies* 10, 1035-1064.

Odders-White, Elizabeth R., 2000. On the Occurrence and Consequences of Inaccurate Trade Classification, *Journal of Financial Markets*, III, 259-286.

Parkinson, M., 1980. The extreme value method for estimating the variance of the rate of return, *Journal of Business* 53, 61-65.

Ravi, Rahul, 2006. "Opacity of a firm and information asymmetry in the financial market". Available at SSRN: http://ssrn.com/abstract=877765

Ritter, Jay R. and Ivo Welch, 2002. A Review Of IPO Activity, Pricing, And Allocations, *Journal of Finance*, 57, 1795-1828.

Ross, S., 1989. Information and volatility: the no-arbitrage martingale approach to timing and resolution irrelevancy. *Journal of Finance* 44 1-17.

Seppi, D. J., 1990. Equilibrium Block Trading and Asymmetric Information, *Journal of Finance*, 45, 73-94.

Seyhun, Nejat, 1986. Insiders' Profits, Costs of Trading, and Market Efficiency, *Journal of Financial Economics* 16, 189-212.

Smith, C.W., and Watts, R.L., 1992. The investment opportunity set and corporate financing, dividend, and compensation policies. *Journal of Financial Economics*, 32(3), 263-292.

Yosha, O., 1995. Information disclosure costs, and the choice of financing source. *Journal of Financial Intermediation*. 4, 3-20.

Zeghal, Daniel, 1984. Firm Size and the Informational Content of Financial Statements, *Journal of Financial and Quantitative Analysis* 19, 299-310.

Appendix

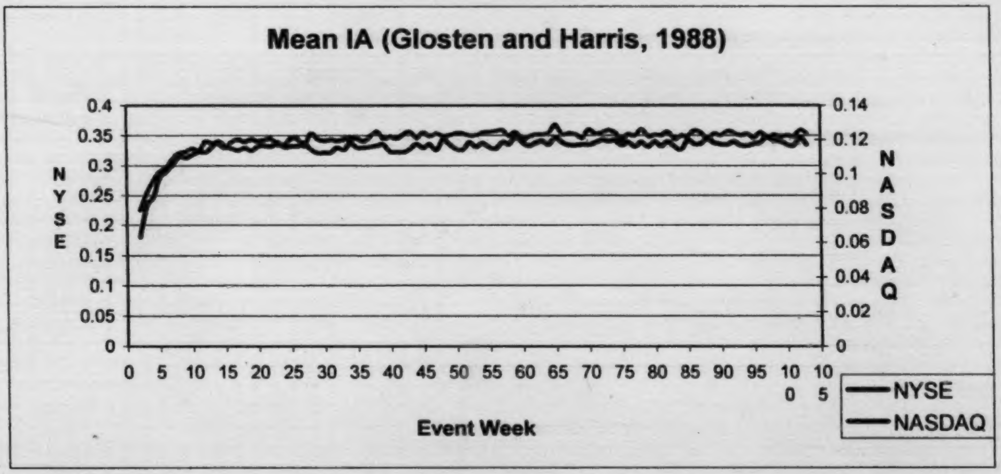

Figure 2.1(A): **The evolution of post-IPO adverse selection cost of trading.**

The graphs trace the mean information asymmetry (the Glosten and Harris measure of information asymmetry) as a function of firm age. The vertical axis measures the Glosten and Harris adverse selection parameter, in the week (IA). The horizontal axis represents the number of weeks post-IPO.

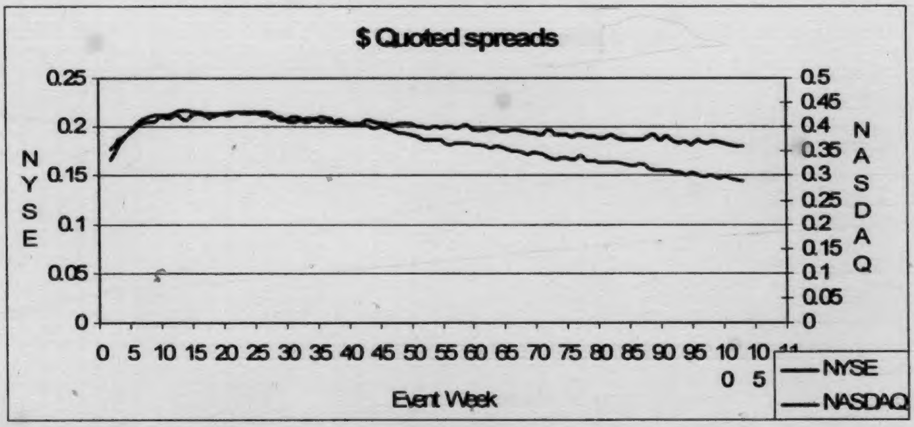

Figure 2.1(B):The evolution of post-IPO quoted spreads.

The figure presents the average quoted spread in each event week. The results for the NYSE and NASDAQ are shown separately. The vertical axis measures the average quoted spread. The horizontal axis represents the post-IPO age of the firm.

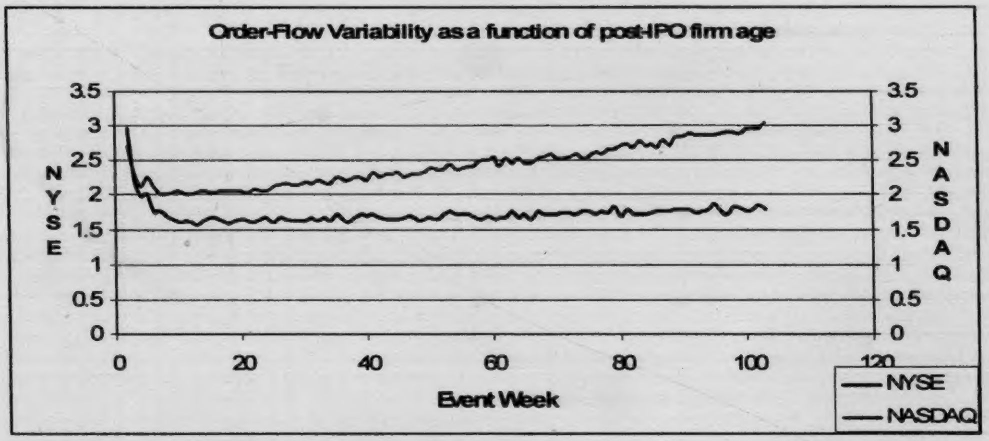

Figure 2.2: Order-flow variability.

The figure presents the cross-sectional mean of the standard deviation of signed order-flow $\left(\sigma_{OF,t}\right)$. In order to compute this variable, we divide trading time in week t into 15-minute intervals. The order-flow variability for stock i in week t $\left(\sigma_{OF,i,t}\right)$ is computed as the standard deviation of the 15-minute net order-flow within the week.

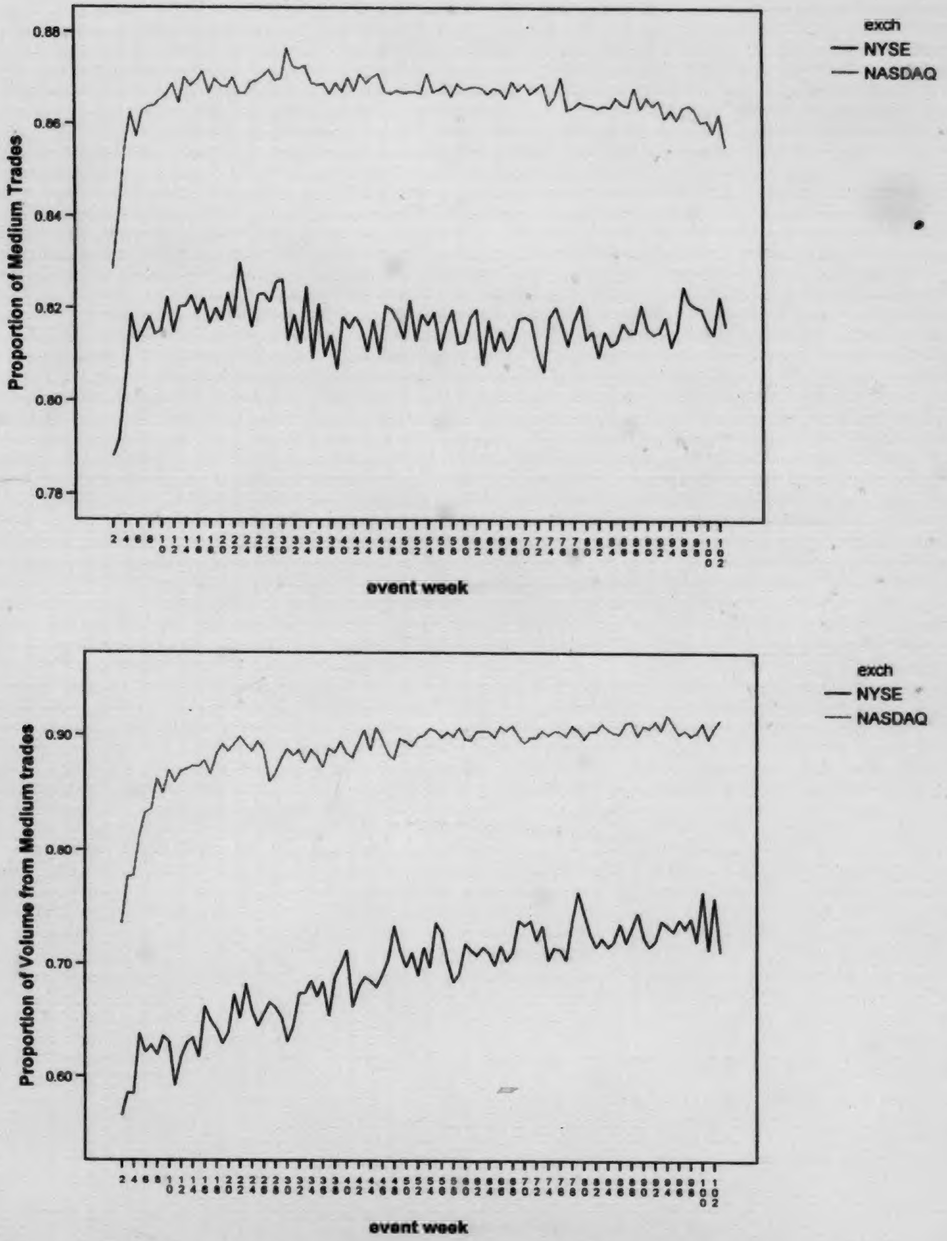

Figure 2.3: **The proportion of medium trades in the IPO-aftermarket.**

Medium trades are trades of 500 to 9900 shares. The figure presents the proportion of medium trades and the proportion of volume occurring due to medium trades in each event week. The x-axis denotes the post-IPO age of the firm (in weeks).

Table 2.1: Pair-wise correlation between measures of information asymmetry.

The measures are Glosten and Harris, 1988 (GH), George Kaul and Nimalendaran, 1991 (GKN), Lin, Sanger and Booth, 1995 (LSB) and Hasbrouck, 1991 (HBRK).
The correlations are computed for each firm in the sample using the time-series of weekly observations for the five measures. The table reports the cross-sectional average of the firm-level correlations, separately for NYSE/AMEX and NASDAQ stocks.

Exchange			GH	GKN	LSB	HBRK
NYSE	GH	Pearson Correlation	1	.915	.928	.777
		Sig. (2-tailed)	.	.000	.000	.000
		N	306	306	306	306
	GKN	Pearson Correlation	.915	1	.876	.692
		Sig. (2-tailed)	.000	.	.000	.000
		N	306	306	306	306
	LSB	Pearson Correlation	.928	.876	1	.834
		Sig. (2-tailed)	.000	.000	.	.000
		N	306	306	306	306
	HBRK	Pearson Correlation	.777	.692	.834	1
		Sig. (2-tailed)	.000	.000	.000	.
		N	306	306	306	306
NASDAQ	GH	Pearson Correlation	1	.760	.899	.636
		Sig. (2-tailed)	.	.000	.000	.000
		N	1574	1574	1574	1574
	GKN	Pearson Correlation	1	1	.746	.593
		Sig. (2-tailed)	0	.	.000	.000
		N	1574	1574	1574	1574
	LSB	Pearson Correlation	.899	.746	1	1
		Sig. (2-tailed)	.000	.000	.	0
		N	1574	1574	1574	1574
	HBRK	Pearson Correlation	1	.593	.573	1
		Sig. (2-tailed)	0	.000	.000	.
		N	1574	1574	1574	1574

Table 2.2(A): Summary firm characteristics for NYSE and NASDAQ listed IPOs.

The sample consists of 289 IPOs on the NYSE and 1661 IPOs on NASDAQ from January 1993 to December 1998. Age is the number of days from inception through the IPO date. The p-value tests the equality of the NYSE and NASDAQ means.

	NYSE		NASDAQ		All IPOs		p-values
	Mean	Std Deviation	Mean	Std Deviation	Mean	Std Deviation	
Number of IPOs	289		1661		1950		
Market Cap (Mil. $)	700.57	1594.12	183.81	259.24	260.39	682.99	0.000
Age (days)	2736	4545	2048	2835	2165	3196	0.167
Number of employees	3612	4755	597	1450	1109	2614	0.000
Book to Market	0.35	0.25	0.36	0.39	0.36	0.37	0.836

Table 2.2(B): Summary offer characteristics of the IPO firms in Table 2.1(A).

Float is the percentage of shares outstanding offered in the IPO. Shareholders provides the percentage of the offered shares sold by the existing pre-IPO shareholders. Company sold the remaining shares. Expenses is the direct costs incurred by the issuer (as a percentage of the capital raised). Filing range is the IPO filing price range. Underwriter Rank is created as described in Carter and Manaster (1990).

	NYSE		NASDAQ		All IPOs		p-values
	Mean	Std Deviation	Mean	Std Deviation	Mean	Std Deviation	
Offer price	$16.50	$4.64	$12.08	$3.70	$12.83	$4.21	0.000
Offer Proceeds (Mil. $)	250.4	570.71	48.98	53.66	83.26	251.09	0.001
Offered Shares (Mill.)	13.81	25.58	3.7	3.53	5.42	11.61	0.000
Float (% Outstanding)	33.14%	17.65%	30.58%	13.17%	31.02%	14.04%	0.114
Shareholders (% Float)	14.83%	28.96%	7.90%	16.89%	9.08%	19.61%	0.032
Company (% Float)	85.09%	29.11%	92.01%	16.98%	90.83%	19.72%	0.030
Expenses (% of capital raised)	2.01%	2.19%	3.02%	1.87%	2.85%	1.96%	0.000
Filing range	14.97%	3.54%	18.20%	4.83%	17.74%	4.80%	0.000

Table 2.3: **Correlation coefficients.**

The table presents the time-series averages of the cross-sectional correlations between the explanatory variables. Panel A presents the averages of the weekly correlations while Panel B presents the average monthly correlations. Here, *Inst* is the number of institutional shareholders; *Size* is the market capitalization of the firm; σ_{OF} is order-flow variability; MB is the market to book ratio; *Analyst* is the number of institutional analysts following the firm and $\sigma_{forecast}$ is the dispersion in their forecasts; σ_{res} is the standard deviation of the market model residual; and PT_{small} PT_{medium} are the proportion of small and medium size trades. Volatility is the natural logarithm of the ratio of the high and low prices during the week.

Panel A: Average weekly cross-sectional correlations

	ln(Size)	ln(Price)	ln(Vol)	σ_{OF}	PT_{small}	PT_{medium}	volatility
ln(Size)	1	0.799	0.523	0.263	0.345	-0.391	-0.086
ln(Price)	0.799	1	0.349	0.23	0.4	-0.42	-0.116
ln(Vol)	0.523	0.349	1	0.176	0.188	-0.228	-0.133
σ_{OF}	0.263	0.23	0.176	1	0.257	-0.223	0.078
PT_{small}	0.345	0.4	0.188	0.257	1	-0.933	-0.023
PT_{medium}	-0.391	-0.42	-0.228	-0.223	-0.933	1	0.038
volatility	-0.086	-0.116	-0.133	0.078	-0.023	0.038	1

Panel B: Average monthly cross-sectional correlations

	ln(Size)	ln(Price)	Ivol	σ_{OF}	PT_{small}	PT_{med}	σ_{res}	MB	ln (Analyst)	ln(Inst)	$\sigma_{forecast}$
ln(Size)	1	0.834	0.622	0.299	0.356	-0.408	-0.083	0.031	0.381	0.532	0.328
ln(Price)	0.834	1	0.384	0.205	0.402	-0.43	-0.139	0.041	0.305	0.364	0.288
Ivol	0.622	0.384	1	0.452	0.15	-0.215	0.048	0.019	0.359	0.407	0.298
σ_{OF}	0.299	0.205	0.452	1	0.821	-0.819	0.024	0.01	0.206	0.25	0.137
PT_{small}	0.356	0.402	0.15	0.821	1	-0.945	-0.048	0.033	0.154	0.115	0.117
PT_{med}	-0.408	-0.43	-0.215	-0.819	-0.945	1	0.067	-0.026	-0.158	-0.148	-0.127
σ_{res}	-0.083	-0.139	0.048	0.024	-0.048	0.067	1	-0.021	-0.018	-0.07	-0.017
MB	0.031	0.041	0.019	0.01	0.033	-0.026	-0.021	1	0.011	0.013	0.008
ln(Analyst)	0.381	0.305	0.359	0.206	0.164	-0.158	-0.018	0.011	1	0.28	0.726
ln(Inst)	0.532	0.364	0.407	0.25	0.115	-0.148	-0.07	0.013	0.28	1	0.206
$\sigma_{forecast}$	0.328	0.288	0.298	0.137	0.117	-0.127	-0.017	0.008	0.726	0.206	1

Table 2.4: **Time-series patterns in information asymmetry.**

The table presents descriptive statistics on information asymmetry (the GH, 1988 measure of information asymmetry). Information asymmetry is computed for each stock and in each event week and the table reports cross-sectional statistics separately for NYSE and NASDAQ stocks.

	NYSE				NASDAQ			
event week	Count	Mean	Median	Std Deviation	Count	Mean	Median	Std Deviation
2	289	0.17997	0.17228	0.09795	1661	0.07881	0.07055	0.06317
3	288	0.23149	0.23114	0.10925	1578	0.08891	0.08053	0.06517
4	291	0.24713	0.24800	0.10575	1602	0.09602	0.09037	0.06363
5	291	0.27947	0.28713	0.10360	1650	0.10136	0.09440	0.06366
6	288	0.28990	0.29871	0.10055	1640	0.10379	0.09306	0.07431
7	290	0.30128	0.31103	0.09619	1635	0.10887	0.09896	0.06914
8	289	0.31525	0.32283	0.09261	1596	0.11224	0.10232	0.07802
9	292	0.31300	0.32730	0.10509	1592	0.11305	0.10161	0.07403
10	290	0.31844	0.32791	0.10079	1592	0.11518	0.10619	0.07270
11	290	0.32276	0.32296	0.10431	1598	0.11360	0.10254	0.07395
12	286	0.32213	0.32929	0.09798	1590	0.11898	0.10623	0.08377
13	291	0.33053	0.33721	0.10616	1553	0.11792	0.10628	0.08282
14	288	0.34021	0.35081	0.10586	1540	0.11782	0.10832	0.07188
15	287	0.33278	0.34387	0.10487	1581	0.11530	0.10580	0.07065
24	285	0.33067	0.34055	0.10310	1489	0.11729	0.10649	0.07645
25	285	0.33248	0.33923	0.10900	1530	0.12065	0.10698	0.08253
26	284	0.33279	0.34461	0.10673	1510	0.12062	0.10989	0.07567
27	284	0.33772	0.35354	0.10523	1518	0.11640	0.10696	0.07689
28	284	0.32713	0.33889	0.10855	1493	0.12364	0.11227	0.08256
51	286	0.32635	0.33556	0.11167	1417	0.12377	0.11622	0.07678
52	284	0.33820	0.34434	0.10819	1430	0.12239	0.11122	0.07597
75	278	0.33406	0.34928	0.11423	1324	0.12001	0.10920	0.07240
76	276	0.33885	0.35442	0.09965	1301	0.12320	0.11348	0.07571
104	260	0.33511	0.34382	0.11586	1186	0.12386	0.11548	0.07172

Table 2.5: **Time-series analysis of weekly information asymmetry.**

The table presents the cross-sectional average coefficients from the following time-series regression:

$$IA_{i,t} = \alpha_i + \beta_{1,i} \times IA_{i,t-1} + \beta_{2,i} \times IA_{mkt,t-1} + \beta_{3,i} \times \sigma_{OF,i,t-1} + \beta_{4,i} \times PT_{small,i,t-1} + \beta_{5,i} \times \ln\left(Price_{i,t-1}\right) +$$

$$\beta_{6,i} \times \ln\left(Vol_{i,t-1}\right) + \beta_{7,i} \times Volatility_{i,t-1} + \varepsilon_i$$

The model is estimated separately for each of the firms 'i' in the sample. R^2 represent the average adjusted R^2. Standard errors are clustered by calendar time.

Panel A: NYSE (Week 2 to week 52 post-IPO)

	1	2	3	4	5	6	7	8	9	10	11
(Constant)	0.31703***	0.37164***	0.32974***	0.32069***	0.11507	0.08045*	0.33314***	0.35138***	0.40081***	0.47114***	0.37039***
Lag(gh)					0.17912***	0.19014***	0.18643***	0.11893***	0.11469***	0.1251***	0.06284***
IA_{mkt}	0.03581**					0.00628*	0.11529	0.01035*	0.02119*	0.00921*	0.0167*
Lag(ln_Price)					0.046**	0.05965***	0.00527*	0.02218*	0.03667***	0.00146*	0.04628***
Lag(ln_vol)							-0.00974*	-0.0103***	-0.01761***	-0.0172***	-0.01536***
Lag(σ(OF))		-0.02617***						-0.01047***			-0.01506***
Lag(PT_{small})			-0.02512*						-0.10373***		-0.09135***
Lag(volatility)				0.05264*						0.06701*	0.16684***
R^2	0.011	0.038	0.021	0.017	0.11	0.131	0.179	0.207	0.205	0.214	0.271

Panel B: NASDAQ (Week 2 to week 52 post-IPO)

NASDAQ											
(Constant)	0.12032***	0.13419***	0.12904***	0.11948***	-0.08983*	0.12273	0.19248***	0.23021***	0.20564***	0.22136***	0.20977***
Lag(gh)					0.00423*	0.00629*	0.00866*	0.0064*	0.01358*	0.0057*	0.03164***
IA_{mkt}	0.01972*					0.29306	0.022*	0.01495*	0.04931*	0.06886*	0.01422*
Lag(ln_Price)					0.07673*	0.0116*	0.00531*	0.02103***	0.00032*	0.01687**	0.0043*
Lag(ln_vol)							-0.00604**	-0.00519***	-0.00658***	-0.00649***	-0.00621***
Lag(σ(OF))		-0.00602*						-0.0034*			-0.0038**
Lag(PT_{small})			-0.04837***						-0.08606***		-0.03232*
Lag(volatility)				-0.00178*						0.01898*	0.04429***
R^2	0.017	0.021	0.024	0.017	0.07	0.105	0.148	0.178	0.186	0.18	0.243

Panel C: NYSE (Week 53 to week 104 post-IPO)

	1	2	3	4	5	6	7	8	9	10	11
(Constant)	0.32569***	0.34166***	0.32758***	0.33188***	-0.15702*	-0.46459*	-0.34121*	-0.23548*	0.41165***	-0.52004*	0.40445***
Lag(gh)					0.09469***	0.10804***	0.1222***	0.09153***	0.11869***	0.09847***	0.05987***
IA_{mkt}	0.05189*					0.06078*	0.06243*	0.08279*	0.03493*	0.21176	0.00541*
Lag(ln_Price)					0.13895	0.23294	0.25864	0.20116	0.02116*	0.34912	0.00129*
Lag(ln_vol)							-0.01435**	-0.0087***	-0.01084**	-0.01556***	-0.00777***
Lag(s(OF))		-0.00007*						-0.01195*			-0.00289*
Lag(PT_{small})			-0.07794						-0.12887*		-0.06333*
Lag(volatility)				0.0425*						0.19707**	0.18296***
R^2	0.01	0.022	0.017	0.015	0.087	0.113	0.15	0.178	0.178	0.177	0.235

Panel D: NASDAQ (Week 53 to week 104 post-IPO)

(Constant)	0.12471***	0.137***	0.13572***	0.12524***	0.28705**	0.19939***	0.25759	0.06742*	0.1386***	0.19566***	0.13831***
Lag(gh)					0.00942*	0.02125**	0.01127*	0.00283*	0.00807*	0.00762*	0.00336*
IA_{mkt}	0.06612*					0.00507*	0.39676*	0.04764*	0.03258*	0.0552*	0.08934*
Lag(ln_Price)					-0.05425*	-0.02317*	-0.05275*	0.00445*	0.00114*	-0.00791*	0.00435*
Lag(ln_vol)							-0.0045*	-0.00461*	-0.00055*	-0.00436***	-0.00224*
Lag(s(OF))		-0.00391*						-0.00392*			-0.00008*
Lag(PT_{small})			-0.07178**						-0.05007*		-0.05696*
Lag(volatility)				0.05184*						0.07473*	0.02967*
R^2	0.016	0.019	0.021	0.018	0.076	0.111	0.148	0.176	0.182	0.178	0.243

Table 2.6: **Difference between information asymmetry for IPOs vs. carve-outs.**

The table presents the coefficients from the following panel regression (Random effects model):

$$IA_{i,t} = \alpha + \beta_1 \times D_{carveout,t} + \beta_2 \times IA_{mkt,t} + \beta_3 \times \sigma_{OF,i,t} + \beta_4 \times PT_{small,i,t} + \beta_5 \times \ln\left(Price_{i,t}\right) +$$
$$\beta_6 \times \ln\left(Vol_{i,t}\right) + \beta_7 \times \ln(Size_{i,t}) + \beta_8 \times Volatility_{i,t} + \varepsilon_{i,t}$$

where: $D_{carveout}$ takes the value 1 if the firm is a carve-out, and zero for a new IPO. IA_{mkt} is the adjusted average information asymmetry across all stocks in the market. Size is market capitalization, Volatility is the natural logarithm of the ratio of the high and low prices during the week, σ_{OF} is the standard deviation of 15-minute signed order-flow within the week, Price is the average transaction price in the week, Vol is the total trading volume in the week, and PT_{medium} is the proportion of medium size trades. Clustered (by calendar time.), Robust standard errors are used for inference.

Panel A: NYSE

NYSE								
$D_{carveout}$	0.004*	0.0065**	0.0065**	0.0006	0.0015	0.0016	0.0012	0.0007
IA_{mkt}								0.0001
ln(Price)		0.0478***	0.0473***	0.0499***	0.0475***	0.053***		0.055***
ln(Size)							-0.0317***	-0.0024
ln(Vol)			-0.0181***	-0.0157***	-0.0193***	-0.0159***	-0.0177***	-0.0158***
σ_{OF}		-0.0146***		-0.0075***		-0.0082***	-0.0068***	-0.0083***
Volatility				0.0522***	0.0489***	0.0509***	0.0417***	0.052***
PT_{medium}			0.0738***			0.0748***	0.0558***	0.0755***
Constant	0.332***	0.2175***	0.3517***	0.3809***	0.4184***	0.3143***	0.0965***	0.3381***
R^2	0.0000	0.0435	0.0901	0.1175	0.1126	0.1247	0.0187	0.1330

Panel B: NASDAQ

NASDAQ								
$D_{carveout}$	0.0082**	0.0075**	0.0059	0.0037	0.0037	0.0038	0.0046*	0.0043
IA_{mkt}								0.0031
ln(Price)		0.008*	0.032***	0.057***	0.052***	0.073***		0.028***
ln(Size)							-0.0068*	-0.0049*
ln(Vol)			-0.0061***	-0.0081***	-0.0089***	-0.0082***	-0.0087***	-0.0086***
σ_{OF}		-0.0005***		-0.0003***		-0.0003***	-0.0003***	-0.0003***
Volatility				0.0515***	0.0508***	0.052***	0.0527***	0.0532***
PT_{medium}			0.0392***			0.0406***	0.0418***	0.0426***
Constant	0.1312***	0.1309***	0.1622***	0.2033***	0.2128***	0.1657***	0.1101***	0.1237***
R^2	0.0004	0.0063	0.0152	0.1374	0.1352	0.1412	0.1434	0.1453

Table 2.7: **Cross sectional determinants of weekly information asymmetry.**

$$IA_{i,t} = \alpha_i + \beta_{1,i} \times \sigma_{OF,i,t-1} + \beta_{2,i} \times PT_{small,i,t-1} + \beta_{3,i} \times PT_{medium,i,t-1} + \beta_{4,i} \times \ln\left(Price_{i,t-1}\right) +$$

$$\beta_{5,i} \times IA_{mkt,t} + \beta_{6,i} \times \ln\left(Vol_{i,t-1}\right) + \beta_{7,i} \times \ln(Size_{i,t-1}) + \beta_{8,i} \times Volatility_{i,t-1} + \varepsilon_i$$

where: Size is market capitalization, Volatility is the natural logarithm of the ratio of the high and low prices during the previous week, σ_{OF} is the standard deviation of 15-minute signed order-flow within the previous week, Price is the average transaction price in the week, Vol is the total trading volume in the week, and PT_{small} PT_{medium} are the proportion of small and medium size trades. IA$_{mkt}$ is the average information asymmetry in the market (adjusted for the dependent firm). Standard errors are clustered by calendar time.

Panel A: NYSE (With Proportion of Small Trades as explanatory variable)

Event Week	(Constant)	ln(Size)	ln(Price)	ln(Volume)	σ_{OF}	IA$_{mkt}$	PT$_{small}$	Volatility	Adj. R^2
3	0.4351***	-0.015*	0.1064***	-0.0179**	-0.0126**	-0.0025	-0.2545***	0.1403***	0.179
4	0.2423***	-0.0008	0.104***	-0.0209***	-0.0274***	0.0548*	-0.1746***	0.937***	0.259
5	0.6144***	-0.02***	0.0921***	-0.0239***	-0.0108	0.0262**	-0.1986***	0.1799***	0.233
6	0.5813***	-0.0263***	0.0813***	-0.0112	-0.0115**	0.0172**	-0.1731**	0.0491	0.149
7	0.567***	-0.0156**	0.078***	-0.0172**	-0.0207**	-0.0037	-0.2393***	0.0639*	0.145
8	0.6961***	-0.0252***	0.0381**	-0.0112	0.0006	0.0068	-0.1654**	-0.0295	0.089
9	0.5759***	-0.0152**	0.0818***	-0.022***	-0.0097*	-0.0048	-0.2217***	0.2104***	0.129
10	0.593***	-0.0288***	0.0774***	-0.0063	-0.0212**	-0.0193	-0.0756	0.0441	0.105
11	0.7541***	-0.0231***	0.0581***	-0.019***	-0.019**	-0.009	-0.1978***	-0.0224	0.167
12	0.6058***	-0.0268***	0.1007***	-0.0162**	-0.0037	-0.0022	-0.256***	0.1409***	0.170
24	0.5547***	-0.0252***	0.0758***	-0.0038	-0.0233***	-0.0012	-0.2744***	0.0679**	0.132
25	0.5876***	-0.0246***	0.0749***	-0.0103	-0.0214**	0.0229	-0.1323*	0.1932***	0.243
26	0.5472***	-0.0185**	0.0722***	-0.0137*	-0.032***	0.1325**	-0.0441	0.1122***	0.247
27	0.6124***	-0.0191**	0.0682***	-0.0134*	-0.0213**	0.004	-0.2059***	0.062**	0.149
28	0.7636***	-0.0317***	0.0762***	-0.0166**	-0.01**	0.0013	-0.222***	0.0619***	0.180
51	0.6956***	-0.0223***	0.0653***	-0.02**	-0.0031*	-0.0029	-0.156*	-0.002	0.096
52	0.6935***	-0.0263***	0.0512***	-0.0134*	-0.0012	-0.0179	-0.0499	0.0939***	0.131
102	0.7054***	-0.0255***	0.0871***	-0.0201***	-0.0097*	0.022**	-0.1783**	0.1167***	0.287
104	0.6594***	-0.0234**	0.0772***	-0.0194**	-0.0008**	0.0314	-0.1073	0.044***	0.126

Panel B: NYSE (With Proportion of Medium Trades as explanatory variable)

Event Week	(Constant)	ln(Size)	ln(Price)	ln(Volume)	σ_{OF}	IA_{mkt}	PT_{medium}	Volatility	Adj. R^2
3	0.1125	-0.0141*	0.1027***	-0.0104	-0.0158***	-0.0026	0.2405***	0.1353***	0.174
4	0.0935	0.0001	0.0908***	-0.014**	-0.0304***	0.0536*	0.0821	0.9309***	0.246
5	0.3879***	-0.0197***	0.0865***	-0.0167**	-0.0149**	0.0263**	0.1535**	0.1758***	0.224
6	0.4188***	-0.0255***	0.0741***	-0.0064	-0.0139**	0.0171*	0.1117	0.0508	0.141
7	0.3231***	-0.0157**	0.0711***	-0.0086	-0.0268***	-0.0041	0.1649**	0.0615*	0.129
8	0.5028***	-0.0256***	0.0361**	-0.0061	-0.0024*	0.0073	0.1506**	-0.0313	0.087
9	0.3125***	-0.0149*	0.0793***	-0.0157**	-0.0148**	-0.0052	0.2044***	0.2036***	0.127
10	0.5189***	-0.0288***	0.0753***	-0.0039	-0.023**	-0.0196	0.0529	0.0433	0.103
11	0.534***	-0.0232***	0.0558***	-0.0131*	-0.0241**	-0.0093	0.1661**	-0.0272	0.163
12	0.3615***	-0.0272***	0.095***	-0.0089	-0.0092	-0.0027	0.183**	0.1382***	0.155
24	0.2465**	-0.0262***	0.0743***	0.0032	-0.0271***	-0.0014	0.2528***	0.0631*	0.130
25	0.459***	-0.0247***	0.0729***	-0.0067	-0.0239***	0.0215	0.0943	0.1912***	0.240
26	0.5306***	-0.0187**	0.07***	-0.0125*	-0.0325***	0.1304**	0.0081	0.1115***	0.246
27	0.4492***	-0.0202***	0.0641***	-0.0076	-0.0246***	0.0044	0.116	0.0608**	0.133
28	0.5745***	-0.0325***	0.0723***	-0.0111	-0.0124*	0.0016	0.1419*	0.0609**	0.167
51	0.5057***	-0.0223***	0.0649***	-0.0164**	-0.0057*	-0.0033	0.1588*	-0.0015	0.097
52	0.6205***	-0.0261***	0.0519***	-0.0125*	-0.002	-0.0183	0.0641	0.0939***	0.132
102	0.4869***	-0.0252***	0.0871***	-0.0169**	-0.0122*	0.022**	0.1868**	0.1155***	0.289
104	0.5169***	-0.0234**	0.0777***	-0.0176**	-0.0017**	0.0314	0.1294	0.043***	0.128

Panel C: NASDAQ (With Proportion of Small Trades as explanatory variable)

Event Week	(Constant)	ln(Size)	ln(Price)	ln(Volume)	σ_{OF}	IA_{mkt}	PT_{small}	Volatility	Adj. R^2
3	0.1731***	0.0003	0.024***	-0.0109***	-0.0006***	0.017	-0.1405***	0.1155***	0.260
4	0.1509***	0.002	0.0153***	-0.0079***	-0.0006***	-0.0002	-0.148***	0.0572***	0.101
5	0.1966***	-0.0016	0.0141***	-0.0078***	-0.0005**	0.0084	-0.1586***	0.0575***	0.091
6	0.1685***	0.0013	0.0203***	-0.0102***	-0.0004**	-0.004	-0.1584***	0.0992***	0.162
7	0.2409***	-0.006**	0.0214***	-0.0094***	-0.0009***	-0.0055	-0.1226***	0.125***	0.176
8	0.1613***	0.0066**	0.0072	-0.0106***	-0.0007**	-0.0079	-0.1803***	0.073***	0.100
9	0.2351***	0.0046	0.0077	-0.0158***	-0.0002*	0.014	-0.1227***	0.0693***	0.119
10	0.2367***	0.005	0.0071	-0.0163***	-0.0001	0.005	-0.1499***	0.1264***	0.181
11	0.2385***	0.0018	0.0129**	-0.0145***	-0.0002*	-0.0048	-0.1441***	0.0974***	0.162
12	0.1689***	0.0079**	0.0025	-0.0121***	-0.0009***	0.0142	-0.1443***	0.096***	0.115
24	0.1915***	0.0085***	0.0006	-0.0147***	-0.0001	0.0041	-0.1192***	0.1118***	0.121
25	0.1817***	0.0047	0.0085*	-0.0118***	-0.0007**	0.0119	-0.1208***	0.1128***	0.170
26	0.2182***	0.0016	0.0133***	-0.0122***	-0.0003*	-0.0023	-0.1537***	0.1006***	0.150
27	0.2093***	0.0039	0.0052	-0.0126***	-0.0002	0.002	-0.1329***	0.1083***	0.193
28	0.1815***	0.0055*	0.0058	-0.0105***	-0.0002	0.0031	-0.2027***	0.0782***	0.126
51	0.2136***	0.0028	0.013***	-0.0126***	-0.0001	-0.0011	-0.1245***	0.0946***	0.187
52	0.1814***	0.0014	0.016***	-0.0099***	-0.0004*	0.0119	-0.0716***	0.0784***	0.116
102	0.2286***	-0.0008	0.0195***	-0.0119***	-0.0004*	-0.003	-0.0502***	0.0921***	0.212
104	0.1872***	0.006*	0.0089**	-0.013***	-0.0004***	-0.0012	-0.0151	0.0532***	0.139

Panel D: NASDAQ (With Proportion of Medium Trades as explanatory variable)

Event Week	(Constant)	ln(Size)	ln(Price)	ln(Volume)	σ_{OF}	IA_{mkt}	PT_{medium}	Volatility	Adj. R^2
3	0.0692**	0.0001	0.0177***	-0.0089***	-0.0009***	0.0146	0.0949***	0.1146***	0.248
4	0.0517	0.0012	0.0091*	-0.0057***	-0.0009***	-0.0006	0.0919***	0.0565***	0.082
5	0.0853***	-0.0026	0.0089*	-0.0053***	-0.0007**	0.0086	0.1014***	0.0558***	0.071
6	0.0779**	0.0001	0.0154***	-0.008***	-0.0007***	-0.0053	0.0837***	0.098***	0.145
7	0.1775***	-0.0072**	0.0173***	-0.0072***	-0.0013***	-0.0059	0.0548**	0.1241***	0.166
8	0.0275	0.0054*	0.0034	-0.0077***	-0.0012**	-0.0082	0.1188***	0.071***	0.084
9	0.1597***	0.004	0.0042	-0.0138***	-0.0006*	0.0134	0.0644***	0.0687***	0.111
10	0.1589***	0.0039	0.0032	-0.0139***	-0.0008**	0.0038	0.0663***	0.1241***	0.168
11	0.1525***	0.0009	0.0091	-0.0119***	-0.0007*	-0.0046	0.0704***	0.0966***	0.150
12	0.0709*	0.0069**	-0.0003	-0.0099***	-0.0014**	0.0144	0.0873***	0.0951***	0.105
24	0.1624***	0.0072**	-0.0032	-0.0124***	-0.0006*	0.005	0.0168	0.1108***	0.110
25	0.1162***	0.0041	0.0059	-0.0102***	-0.0012**	0.0105	0.0552**	0.1114***	0.160
26	0.1012***	0.0006	0.0116**	-0.0099***	-0.0008*	-0.0025	0.103***	0.0984***	0.139
27	0.1013**	0.0031	0.0038	-0.0107***	-0.0007*	0.002	0.0974***	0.1077***	0.189
28	0.0523	0.0039	0.0023	-0.0065***	-0.0012**	0.0024	0.1023***	0.0761***	0.104
51	0.1285***	0.002	0.0119**	-0.011***	-0.0004*	-0.0058	0.0749***	0.0935***	0.180
52	0.14***	0.0008	0.015***	-0.0088***	-0.0005*	0.0112	0.0349	0.0778***	0.111
102	0.1849***	-0.0011	0.0193***	-0.0114***	-0.0004**	-0.0031	0.0406**	0.0918***	0.211
104	0.1774***	0.0059*	0.0087**	-0.0128***	-0.0004***	-0.0013	0.0089	0.0531***	0.138

Table 2.8: **Cross-sectional determinants of monthly information asymmetry.**

$$IA=\beta_0+\beta_1\times\ln(Inst)+\beta_2\times\ln(Size)+\beta_3\times\sigma_{OF}+\beta_4\times MB+\beta_5\times\ln(Analyst)+\beta_6\times\sigma_{res}+\beta_7\times\sigma_{forecast}+\varepsilon$$

where *Inst* is the number of institutional shareholders; *Size* is the market capitalization of the firm; σ_{OF} is order-flow variability; MB is the firm's market to book ratio; *Analyst* is the number of institutional analysts following the firm and $\sigma_{forecast}$ is the dispersion in their forecasts; σ_{res} is the standard deviation of the market model residual. The CRSP value weighted index return is used as the proxy for the market return. Standard errors are clustered by calendar time.

	Months Post IPO							
NYSE & AMEX	**2**	**8**	**14**	**20**	**26**	**32**	**38**	**44**
(Constant)	0.3339	0.5773***	0.7018***	0.4759***	0.4709***	0.5201***	0.4177***	0.5275***
ln(Inst)	0.0178***	-0.0057	-0.0064	-0.012	-0.0168	-0.0131	-0.0095	-0.0124
ln(Size)	-0.015	-0.0122	-0.0219**	-0.0031	-0.0029	-0.0078	-0.0106	-0.0024
σ(OF)	-0.0285***	-0.0389***	-0.017***	-0.0217**	-0.0177***	-0.0142**	-0.0096	-0.0089
MB	0.0026	-0.0036	0.0071	0.002	0.002	0.0042	0.0001	-0.004
ln(Analyst)	-0.0937	-0.0149	-0.0331	-0.0117	-0.0097	-0.0015	0.0234	-0.0388
σ(res)	1.8397***	0.8434***	0.1324***	0.5462***	0.5834***	0.3086***	0.5802***	0.1276***
σ(forecast)	-0.0004	-0.0002	0.0003	0.0002	0.0003	0.0001	-0.0005	0.0006
Adj R^2	0.224	0.159	0.082	0.028	0.065	0.046	0.021	0.007

NASDAQ	**2**	**8**	**14**	**20**	**26**	**32**	**38**	**44**
(Constant)	0.0573	0.212***	0.1915***	0.1447***	0.1347***	0.1636***	0.1748***	0.242***
ln(Inst)	0.0113**	-0.002	-0.0042	0.0015	-0.0024	-0.0015	-0.0038	0.0046
ln(Size)	0.0038	-0.0068	-0.0012	-0.0017	0.0001	-0.002	-0.0004	-0.0059**
σ(OF)	-0.0007	-0.0009	-0.0006	-0.0012***	-0.0008***	-0.0006***	-0.0005***	-0.0007***
mb	-0.0001	0.0009	0.0009	0.0029***	0.002**	0.0000	0.0002	-0.0004
ln(Analyst)	-0.014	-0.0064	-0.0067	-0.0135	-0.0148	-0.0126	-0.006	-0.0047
σ(res)	0.2153*	0.3571*	0.412**	0.1627	0.3366**	0.3018**	0.5264***	0.6056***
σ(forecast)	0.0001	0.0001	0.0001	0.0001	-0.0002	-0.0001	-0.0001	0.0000
Adj R^2	0.011	0.01	0.008	0.023	0.025	0.019	0.029	0.054

Chapter 3

Understanding the second moment of order-flow: Implications for the trading environment

3.1 Introduction

A large and growing body of empirical microstructure research examines the relation between order imbalance, which measures the difference between buying and selling pressure, and stock prices. In this paper, we extend this literature by examining the link between the second moment of order imbalance, measured by the standard deviation of order imbalance, and prices.

Microstructure theory suggests various reasons why the variability of order imbalance will affect prices. Kyle (1985) shows that the extent to which order imbalance (comprising both informed and uninformed orders) will move prices – the price impact parameter – is decreasing in the variability of uninformed orders. Inventory models such as the one devised by Ho and Stoll (1981) show that a market-maker's inventory costs will increase in the variability of order imbalance. Both arguments imply that the variability of order imbalance will affect the level of trading costs (albeit, in different directions) and thereby affect the expected returns (Amihud and Mendelson, 1986; Brennan and Subrahmanyum, 1996). Finally, the literature focusing on differences in opinion (see for e.g., Miller, 1977) suggests that greater heterogeneity in the investor population will, in the presence of short selling constraints, lead to inflated prices and to subsequent corrections. This literature suggests a negative relation between the variability of order imbalance and subsequent returns.

Motivated by these arguments, we study the relation between the standard deviation of 15-minute order imbalance (or SIGOF for short) and stock returns, the bid-ask spread and its components, and trading volume. We use high frequency data to first compute 15-minute order imbalance and then calculate the standard deviation of this series within each month for a sample of 3870 NYSE stocks over the period from January 1993 through December 2003. On average, the sample consists of 1,765 firms in every calendar month.

For every month in the sample period, we partition the sample into SIGOF quintiles. We find that, on average, higher SIGOF leads to a lower per dollar adverse selection cost of trading, a lower inventory cost (per dollar), a lower bid-ask spread and proportional spread, lower risk-adjusted returns, and higher trading volume. The negative relation between lagged order-flow variability and inventory costs is somewhat puzzling, however, in that more variable order-flow results from either greater divergence in opinions or heavier liquidity trading, and order-flow becomes less informative about the true price and the adverse selection risk of the market maker is reduced (Kyle 1985). The positive relation between SIGOF and volume and the negative relation between SIGOF and future returns is consistent with the predictions of the divergence from the opinion literature. We also find that SIGOF is positively associated with other proxies for divergence in opinions, namely, market capitalization, S&P 500 futures open interest, dispersion in analyst forecasts, and the volatility of trading volume.

Lastly, we explore commonality in SIGOF. Several papers study common effects in order imbalance (Hasbrouck and Seppi, 2001; Harford and Kaul, 2005). Our interest in common effects and order-flow variability is motivated by the idea that differences in opinions could be correlated across stocks. We present evidence of significant commonality in order-flow variability, with 83% of the stocks in our sample tending to move in the same direction. Building on the work of Chordia, Roll, and Subrahmanyam (2001), we show that the adverse selection and inventory components of the spread display commonality, and we link this commonality to co-movement in SIGOF. We provide some evidence that the commonality in liquidity and in the adverse selection and inventory components is at least partially determined by the same factors that determine commonality in order-flow variability. Finally, to understand the drivers of SIGOF, we link stock SIGOF to systematic and idiosyncratic variables. We show that SIGOF contains a systematic component, plausibly associated with aggregate divergence in opinions.

The remainder of this paper is organized as follows. Section 3.2 develops our hypothesis. Section 3.3 describes the construction of the key variables. Section 3.4 details the data and the sample. Section 3.5 presents the results and Section 3.6 offers some conclusions.

3.2 Hypotheses Development

Order-flow measures the active side of the trade. Therefore, order-flow variability (SIGOF) is a plausible measure of stock level differences in opinions, since, if order-flow fluctuates over a day or a month (15 minutes, in our tests), investors likely do not agree. This section develops several hypothesis designed to test the ability of SIGOF to measure stock level and market-wide divergence in opinions and to explore the effects of changes in SIGOF.

3.2.1 SIGOF and trading costs

Models such as the one devised by Kyle (1985) demonstrate that trading costs increase in the degree of the potential information asymmetry between the market-maker and the informed investors. Nevertheless, *ceteris paribus,* trading costs should decline in the variability of uninformed trading as they allow the informed trader to hide trades more effectively.

> H1: At any given point in time, the adverse selection cost of trading should be negatively related to the variability of order-flow.

Ho and Stoll (1981) suggest that asynchronous timing in buy and sell orders imposes inventory management cost on the market-maker. Therefore, a market-maker's inventory costs should increase in the variability of order imbalances.

> H2: At any given point in time, the inventory holding cost of the market-maker should be positively associated with the variability of order-flow.

3.2.2 SIGOF and trading volume

Why do investors trade such enormous quantities? Differences in information alone cannot explain high levels of trading volume (Milgrom and Stokey, 1982). Harris and Raviv (1993) and Kandel and Pearson (1995) show that differences in opinions help to explain the high levels of trading volume, and that a greater divergence in opinion leads to higher trading volume. These differences can arise either due to differences in prior beliefs or due to differences in the way investors interpret public information.

H3: If SIGOF is a measure of divergence in opinions, trading volume should be positively correlated with contemporaneous order-flow variability.

Anshuman, Chordia, and Subrahmaniam (2005) explore the properties of variability in trading volume (σ_{Vol}), and suggest that σ_{Vol} can be interpreted as a measure of divergence in opinions.

H4: If SIGOF is a measure of divergence in opinions, variability in trading volume should be positively correlated with contemporaneous order-flow variability.

3.2.4 SIGOF and dispersion in analyst forecasts

Diether, Malloy, and Scherbina (2002) use dispersion in analyst annual earnings forecasts as a proxy for differences in opinions. They find that stocks with higher dispersion in analysts' earnings forecasts earn significantly lower future returns than do otherwise similar stocks. If order-flow variability is a measure of differences in opinions, we should see a positive contemporaneous relation between SIGOF and dispersion in analyst forecasts.

H5: Dispersion in analyst forecasts should be positively correlated with contemporaneous order-flow variability.

3.2.5 SIGOF and returns

Miller (1977) suggests that short-sale constraints prevent pessimistic opinions from being fully reflected in stock prices. Miller argues that a stock's price will reflect the valuations of optimistic investors because pessimists cannot participate in the market when short sale constraints are in place.[13] Thus, in the presence of short sale constraints, stocks may become overpriced during periods of high differences of opinions about their prospects. Therefore, to the extent that SIGOF is a measure of divergence in opinions, we expect to find a positive contemporaneous and negative lagged relation between SIGOF and stock returns.

H6: Order-flow variability should be positively associated with contemporaneous returns.

H7: Order-flow variability should be negatively associated with returns in the following period.

3.2.6 SIGOF and market-wide divergence in opinions

Miller (1977) convincingly argues that the divergence in opinions is not entirely idiosyncratic, but is correlated with both the systematic and the non-systematic components of a stock's return. While the above set of hypotheses relates to idiosyncratic differences in opinions, the next hypothesis explores the relation between SIGOF and systematic divergence in opinions. Bessembinder, Chan, and Seguin (1996) propose a useful proxy for systematic dispersion in opinions, suggesting that open interest in the

---

[13] This argument is reinforced by the fact that arbitrage is risky and costly (e.g., Pontiff, 1996).

S&P 500 index futures contract captures the cross-sectional dispersion in traders' opinions regarding the market-wide prospects. We accordingly relate the time-series of SIGOF of each stock in the sample to open interest in the S&P 500 index futures contract.

> H8: On average, SIGOF should be positively related to open interest in the S&P 500 index futures contract.

3.2.7 *Co-movement in SIGOF and liquidity*

The final section in our analysis of SIGOF attempts to explore the levels of co-movement in SIGOF. Drawing upon the arguments leading to Hypothesis 8, the systematic component in SIGOF should induce co-movement in order-flow variability across stocks. We take this argument further by suggesting that if the spread and the components of the spread are related to order-flow variability (H1 and H2), then co-movement in SIGOF is likely to induce co-movement in the spread and in its adverse selection and/or in the inventory components.

> H9: Co-movement in order-flow variability will induce co-movement in the spread, the adverse selection component and the inventory component of the spread.

3.3 Construction of Variables and Empirical Methods

We carry out the empirical analysis in three stages. We start with a firm-level contemporaneous analysis to study the association between SIGOF and the adverse selection cost per dollar traded (DVIA), the inventory cost per dollar traded (DVINV), risk-adjusted stock returns, trading volume (vol), market capitalization (Size), variability of trading volume (SigVol), dispersion in analysts' forecasts (DISP), and the number of analysts following a stock (ANAL). In the second stage, we explore the determinants and the effects of SIGOF. Finally, we examine co-movement in SIGOF and its implications for co-movement in liquidity.

We start by classifying all sample trades as either buyer- or seller-initiated. Following the procedure of Lee and Ready (1991), trades are classified as buyer- or seller-initiated if the transaction price is closer to the ask (bid) price of the prevailing quote. The quote must be at least five seconds old. If the trade is exactly at the midpoint of the quote, the "tick test" is employed. In this case, a trade is classified as a buy if the most recent non-zero price change is positive, and classified as a sell if the most recent non-zero price change is negative. Since the trade direction must be inferred, inevitably, some assignment error may occur. However, as shown by Lee and Radhakrishna (2000) and by Odders-White (2000), the Lee Ready (1991) algorithm is accurate.

3.3.1 Measuring order-flow variability

We divide each trading day (9:30 a.m. to 4:00 p.m.) in a given month into 26 15-minute intervals. For every stock in our sample, we compute order-flow (the number of buyer-initiated trades minus the number of seller-initiated trades) in each interval. For a given stock x, $SIGOF_{x,t}$ is the standard deviation of the 15-minute order-flow series in month t. We compute two additional measures of order-flow variability based on alternative definitions of order-flow: first, as the difference between the volume of buyer-initiated trades and the volume of seller-initiated trades; and second, as the difference between the value of buyer-initiated trades and the value of seller-initiated trades. The results are identical, and therefore, for the sake of brevity, this paper only discusses the results corresponding to the number of trades-based measures of order-flow. Moreover, while the volume-based measure is likely to be contaminated by trade size effects, the value measure is likely to be affected by prices. Since several of the hypothesized variables are functions of either price or volume, we believe the choice of number of trades-based measure of order-flow is the most conservative.

3.3.2 Measures of liquidity

We use the quoted spread and proportional quoted spread as two related measures of liquidity. The quoted spread (QSPR) is defined as $QSPR = (P_A - P_B)$ where P_A is the ask price and P_B is the bid price. Defining the quote midpoint as $P_M = (P_A + P_B)/2$, the proportional quoted spread is defined as $PQSPR = (P_A - P_B)/P_M$.

3.3.3 The adverse selection and inventory cost components of the spread

We estimate the components of the bid-ask spread using the method advocated by Lin, Sanger, and Booth (LSB, 1995).[14] This method is based on the approach described in Stoll (1989) and related to the approach used by Huang and Stoll (1997). LSB use a regression approach to estimate the proportion of the effective spread that can be attributed to information asymmetry. The basic idea is that the quote revision reflects the adverse selection component of the spread, while the change in the transaction price reflects the order processing costs and bid-ask bounce.

---

[14] We have also run our analysis using the adverse selection components proposed by Glosten and Harris (1988) and Neal and Wheatley (1998). Our results are robust to the method selected. For the sake of brevity, we only report the results corresponding to LSB (1995).

In the LSB model, information revealed by the trade at time t is reflected in the quote revisions. If P_t is the transaction price at time t, and Q_t is the quote midpoint at time t, then $B_t = B_{t-1} + \lambda S_{t-1}$ and $A_t = A_{t-1} + \lambda S_{t-1}$, where B_{t-1} and A_{t-1} are the prevailing bid and the ask prices at time t. λ can be interpreted as the proportion of the effective spread due to adverse selection. $S_{t-1} = P_{t-1} - Q_{t-1}$ is one-half of the effective spread. The revision in the quote mid point is expressed as

$$\Delta Q_t = \lambda S_{t-1} + \varepsilon_t \qquad \text{......................} (3.1)$$

$$S_t = \theta S_{t-1} + \eta_t \qquad \text{......................} (3.2)$$

where $\Delta Q_t = Q_t - Q_{t-1}$ and $Q_t = \frac{(B_t + A_t)}{2}$. θ represents the order processing cost component of the spread, and $(1 - \lambda - \theta)$ represents the inventory component of the bid-ask spread. We calculate the per dollar adverse selection cost of trading (DVIA) by multiplying λ by the average monthly effective spread and dividing it by the average transaction price for the month. We use the same method to calculate the per dollar inventory cost of trading (DVINV).

3.3.4 Other variables

Trading volume (VOL) is the total number of shares traded in each month, as reported in the Center for Research in Security Prices (CRSP) database. The standard deviation of trading volume (SigVol) is estimated in a manner analogous to the SIGOF measure. We define SigVol as the standard deviation of 15-minute trading volume, calculated across all 15-minute intervals in a given month.

Monthly holding period returns (r) are obtained directly from the CRSP monthly tapes. We calculate the risk-adjusted return using the four-factor model of Carhart (1997). This model is an extension of the Fama and French (1993) three-factor model, incorporating an additional momentum factor. For each month, we run the following regression for firms with more than 17 daily return observations within that month,

$$R_{i,t,d} = \alpha_{i,t} + \beta_{i,m} \times r_{m,d} + \beta_{i,SMB} \times SMB_{t,d} + \beta_{i,HML} \times HML_{t,d} + \beta_{i,MOM} \times MOM_{t,d} + \varepsilon_{i,t,d} \quad (3.3)$$

where, for day d in month t, $R_{i,t,d}$ is stock i's excess return; $r_{m,d}$ is the excess return on the market portfolio; and $SMB_{t,d}$ and $HML_{t,d}$ are the Fama-French (1993) size and book-to-market portfolios. $Mom_{t,d}$ is the momentum factor in month t. $\varepsilon_{i,t,d}$ is the residual with respect to the described factor model. The data for the three factors (HML, SMB, Mom) are obtained from Ken French's website.[15] The risk-adjusted return for stock p in month t is calculated as $\alpha_{i,t}$. The daily risk-adjusted return is given by $r_{i,t,d} = \alpha_{i,t} + \varepsilon_{i,t,d}$.

The market-to-book-ratio of the firm (MB) is calculated as:

$$MB = \frac{(Common\ shares\ outs\tan ding) \times (Share\ \Pr ice) + (Total\ assets) - (Common\ equity)}{(Total\ assets)}$$

where common shares outstanding is obtained as Compustat data # 61. Data # 14 provides the closing share price, and total assets corresponds to data # 44. Common equity corresponds to data # 59. All data item numbers correspond to the Compustat quarterly file.

---

[15] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/

Several existing studies have related firm size (Size) to information production. Size is also inversely related to stock returns. Size is defined as the month-end shares outstanding times the month-end closing price. Dispersion in analysts' forecasts (DISP) is used as a proxy for divergence in opinion and is measured as the standard deviation of current fiscal year earnings forecasts, divided by the consensus mean of current fiscal year earnings forecasts. The DISP data are obtained from I/B/E/S.

We use the number of analysts providing forecasts (ANAL) as a control variable, since, although one expects ANAL to be related to SIGOF, the nature of the relation is an empirical issue. By controlling for the dispersion in analysts' forecasts, the number of analysts following a stock should be positively associated with firm transparency. For opaque firms, an increase in transparency should provide an informational advantage to smarter traders, thereby leading to an increase in divergence in opinions. However, for transparent firms an increase in transparency reduces the smart traders' advantage and thus leads to a reduction in divergence in opinions (Ravi, 2006). Therefore, *ceteris paribus,* the relation between the number of analysts providing earnings forecasts and divergence in opinions should depend on the level of transparency for the sample firms.

The variables for inclusion in exploring the determinants of a stock's order-flow variability are discussed in Section 3.2. Hypotheses 1 through 5 refer to cross-sectional associations involving order-flow variability. To test these hypotheses, we use the following cross-sectional regression model:

$$\begin{aligned} SIGOF_{i,t} = {} & \alpha_t + \beta_{1,t} \times \ln\left(\sigma_{VOL,i,t}\right) + \beta_{2,t} \times \ln\left(Vol_{i,t}\right) + \beta_{3,t} \times \ln\left(Size_{i,t}\right) + \beta_{4,t} \times IA_{i,t} + \\ & + \beta_{5,t} \times INV_{i,t} + \beta_{6,t} \times \ln\left(Anal_{i,t}\right) + \beta_{7,t} \times DISP_{i,t} + \beta_{8,t} \times MB_{i,t} + \beta_{9,t} \times ret_{i,t} + \varepsilon_{i,t} \quad \ldots(3.4) \end{aligned}$$

where the subscripts (i,t) denote stock i and month t. Size corresponds to the market capitalization of the firm; Vol is monthly trading volume; and σ_{Vol} is the variability in 15-minute trading volume. ANAL is the number of analysts following the firm. DISP is the dispersion in analyst forecasts. MB is the market-to-book ratio of the firm. The risk-adjusted contemporaneous return is added to the model as a control variable.

Hypotheses 6 and 8 refer to time-series associations involving SIGOF. To test these, we require the use of a time-series regression model. Several of the variables used in this analysis are persistent through time, and thus raises concerns about the inference of causality and contemporaneous associations. We get around this problem by following Chordia, Roll, and Subrahmanyam (2001) in using monthly proportional changes in the variables, rather than levels in the regression analysis. For example, for the variable M, the proportional change is defined as $(M_t - M_{t-1})/M_{t-1}$. The model used is as follows:

$$\begin{aligned} DSIGOF_{i,t} = {} & \alpha_i + \beta_{1,i} \times DNOIC_t + \beta_{2,i} \times DSize_{i,t} + \beta_{3,i} \times DVol_{i,t} + \beta_{4,i} \times DSigVol_{i,t} \\ & + \beta_{5,i} \times r_{i,t} + \beta_{6,i} \times DANAL_{i,t} + \beta_{7,i} \times DDISP_{i,t} + \beta_{8,i} \times DMB_{i,t} + \varepsilon_{i,t} \qquad \text{.... (3.5)} \end{aligned}$$

The subscripts (i,t) refer to stock i and month t, respectively. D denotes proportional change and the subscript t indicates that the change is being calculated between trading months t-1 and t. NOIC is S&P 500 futures open interest (measured as number of contracts).

3.3.5 Co-movement in SIGOF

Hypothesis 9 addresses whether or not order-flow variability co-moves, and its implications for co-movement in the spread and its components. Two related methods are used to test this hypothesis. First, we use pair-wise correlation analysis. We estimate the pair-wise correlation between the quoted spread for each of the 3,870 firms in the sample $\left(Corr\left(QSPR_{i,t}, QSPR_{j,t}\right)\right)$ for all $i \neq j$. To assess the role of SIGOF, we orthogonalize the quoted spread for firm i with respect to its SIGOF:

$$QSPR_{i,t} = \alpha_i + \beta_i \times SIGOF_{i,t} + \varepsilon_{i,t} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.6)$$

The residuals $\varepsilon_{i,t}$ represent the spread for firm i in month t, while controlling for the order-flow volatility of firm i. We re-estimate the pair-wise residual correlation $Corr\left(\varepsilon_{i,t}, \varepsilon_{j,t}\right)$. Comparing $Corr\left(QSPR_{i,t}, QSPR_{j,t}\right)$ with $Corr\left(\varepsilon_{i,t}, \varepsilon_{j,t}\right)$ provides a way of quantifying the contribution of SIGOF to spread co-movement. We repeat the analysis using proportional spreads as a related measure of liquidity.

The second method for exploring co-movement in liquidity and the role of SIGOF in such co-movement is closely related to the work of Chordia, Roll, and Subrahmanyam (2001). We estimate time-series regressions relating monthly proportional changes in SIGOF for individual stocks to market-wide average order-flow variability, i.e.,

$$DSIGOF_{i,t} = \alpha_i + \beta_i \times DSIGOF_{M,t} + \gamma_{i,1} r_{m,t-1} + \gamma_{i,2} r_{m,t} + \gamma_{i,3} r_{m,t+1} + \gamma_{i,4} \ln\left(\frac{P_{i,t,\max}}{P_{i,t,\min}}\right) + \varepsilon_{i,t} \quad \ldots (3.7)$$

where $DSIGOF_{i,t}$ is the proportional change in order-flow variability for stock i from trading month t-1 to t. $DSIGOF_{M,t}$ is the corresponding change in market-wide SIGOF, calculated as:

$$SIGOF_{M,t} = \frac{\sum_{i=1}^{n} SIGOF_{i,t}}{n}$$

$$DSIGOF_{M,t} = \frac{\left(SIGOF_{M,t} - SIGOF_{M,t-1}\right)}{SIGOF_{M,t-1}}$$

$r_{m,t+1}$, $r_{m,t}$, and $r_{m,t-1}$ are the lead, contemporaneous, and lag market returns, respectively, included in the model to control for any possible effects of returns on order-flow variability. The contemporaneous natural logarithm of the ratio of the maximum and minimum prices of stock i in month t is included as a control for volatility. The β_i coefficients may be interpreted as a measure of co-movement in SIGOF.

In computing the market index $DSIGOF_{M,t}$, the value of stock i is excluded, and thus, the explanatory variable in the above regression is slightly different for each stock's time-series regression. We estimate model (8) for the two measures of liquidity (QSPR and PQSPR), the adverse selection cost per dollar traded (DVIA), and the inventory cost per dollar traded (DVINV).

To explore the role of SIGOF in liquidity co-movement, we control for the effect of SIGOF on QSPR, PQSPR, DVIA, and DVINV, using the OLS specification:

$$M_{i,t} = a_i + b_i SIGOF_{i,t} + \varepsilon_{i,t} \quad \text{......................} \ (3.8)$$

We then repeat the co-movement analysis in Equation (3.7) on $\varepsilon_{i,t}$.

3.4 Sample Selection and Sample Characteristics

The sample period runs from January 1993 to December 2003. Data was retrieved from the NYSE Trade and Quote (TAQ), Compustat, and Center for Research in Security

Prices (CRSP) databases. Analyst data is obtained from the I/B/E/S database. Utilities (SIC code 49 to 50) and firms from the financial sector (SIC code 60 to 68) were excluded because these are regulated industries. ADRs, other securities incorporated outside the US, as well as preferred stocks and other non-common stocks, were excluded.[16] We delete all non-NYSE firms from the sample.[17]

Several filters were employed to ensure the validity of the TAQ data.[18] The first trade of each day is dropped from the analysis, since it usually occurs through a call auction. The TAQ database does not eliminate auto-quotes (passive quotes by secondary market dealers), which may cause the quoted spreads to be artificially inflated. Since no reliable method can exclude auto-quotes in TAQ, only BBO (best bid or offer) eligible primary market (NYSE) quotes were used (Chordia, Roll, and Subrahmanyam 2001, 2002).[19] Following Lee and Ready (1991), any quote delivered less than five seconds prior to the trade is ignored, and the first quote that is at least five seconds prior to the trade is retained.

Order-flow variability $(SIGOF)$ and the adverse selection (DVIA) and the inventory (DVINV) components of the spread are generated from the TAQ data. Our sample period extends from January 1993 through December 2003. The sample size ranges from a minimum of 1,605 firms in January 1993 to a maximum of 2,194 firms in April 1998, and to 2,072 in December 2003. The full sample consists of a total of 747,366,091 matched trade and quote pairs. Using these matched pairs, we compute monthly DVIA, DVINV, and *SIGOF* for each firm in the sample period. We merge these monthly series with monthly volume, size, and return data from the CRSP; quarterly market-to-book is from Compustat; and the number of analysts and the dispersion in analyst earnings forecasts is from I/B/E/S. Our final dataset consists of 162,130 firm-months of data.

---

[16] Securities with CRSP share codes different from 10 or 11 were excluded.

[17] The spread decomposition methodologies used in this paper are appropriate for a specialist market (NYSE), as opposed to dealer markets (NASDAQ). In addition, interpretation of the spread components for NASDAQ trade and quotes is potentially problematic due to the presence of inter-dealer trades in the data. These non-informational trades cannot be identified in the database. Restricting this study to NYSE-based firms also abstracts from differences in market structure.

[18] We drop all trades with a correction indicator other than 0 or 1, and retain only those trades for which the condition is B, J, K, or S. We also drop all trades with non-positive trade size or price. Finally, we omit all trades recorded before opening time or after the closing time of the market. Negative bid-ask spreads and transaction prices are also eliminated. In addition, we eliminate all quotes for which the quoted spread is greater than 20% of the quote midpoint when the quote midpoint is greater than $10, or for which the quoted spread is greater than $2 when the quote midpoint is less than $10. We also eliminate all quotes for which either the ask or the bid moves by more than 50%.

[19] All quotes with condition 5, 7, 8, 9, 11, 13, 14, 15, 16, 17, 19, 20, 27, 28, 29 were excluded.

3.5 Results

Table 3.1 presents the distribution of firms over the sample period and descriptive information. The size of the average firm in the sample increases from 1993 to 2000, and then drops for the rest of the study period. The average market-to-book ratio remains stable and the average number of analysts following a firm declines over the sample period. Among the microstructure variables, the average adverse selection cost of trading, expressed as a percentage of the quoted spread, shows a marginally increasing trend from 1993 to 1998 and then stabilizes for the rest of the sample period. Average trading volume increases almost four-fold between 1993 and 2003. The mean inventory cost component of the spread shows an almost monotonic decline, with a minimum of 12.17% in 2002.

3.5.1 Exploring order-flow variability

Table 3.2 presents the time-series distribution of the mean SIGOF across industries.[20] While the exact ranks of industries with respect to average SIGOF changes from year to year, high tech industries such as health care, drugs and genetic engineering and computer manufacturing have more volatile SIGOF, while industries such as wholesale and construction display stable order-flow. Given that the firms in the former group of industries are more difficult to value accurately, a higher level of dispersion in opinions about their value is likely. The firms in the latter set are high tangible-asset firms, which would be relatively easier to value. The patterns in Table 3.2 provide preliminary support for our interpretation of SIGOF as a measure of divergence in opinions or the level of uninformed trading in the market.

Table 3.3 shows the time-series and cross-sectional variation in average SIGOF, sorting at the end of each month by firm size (Panel A), market-to-book ratio (panel B), and the number of analysts providing earnings forecasts for the firm (panel C). Average SIGOF increases as each of the three firm characteristics increases. The increasing mean across the firm characteristics quintiles are consistent across years. This result supports the dispersion in opinions interpretation of SIGOF.

Average SIGOF is fairly stable until 1996, and increases monotonically from 1996 to 2003 (Table 3.1). Table 3.3 suggests that this time-series pattern is most prominent for the largest 20% of the firms and relatively weak among the smallest 20% of the firms. The same time-series pattern in SIGOF can also be seen with respect to the number of analysts following the firm (ANAL). Given that the correlation between ANAL and size is 0.784 (Table 3.5), this result is not surprising. The time-series pattern in SIGOF is also visible across market-to-book quintiles, though in this case, it is more prominent among the higher quintiles.

---

[20] We use an adapted version of the 14-industry classification, as proposed in Ritter and Welch (2002).

The results in Table 3.3 suggest that larger firms, firms with more growth options, and firms followed by more analysts, tend to have high order-flow variability. To the extent that these characteristics are not orthogonal to each other, these results need to be interpreted with caution. Table 3.4 attempts to further explore the results in Table 3 by examining SIGOF for two-way sort. We divide the sample into size and MB quintiles (Panel A), MB and ANAL quintiles (Panel B), and size and ANAL quintiles (Panel C).

Controlling for size, the book-to-market effect observed in Table 3.3 becomes much weaker. Order-flow variability is low for smaller firms and high for larger firms. Similarly, when controlling for the number of analysts providing earnings forecasts, the market-to-book effect becomes, once again, considerably weaker. A possible explanation for the weakening of the market-to-book effect is the ambiguous nature of the variable. While high market-to-book is usually interpreted as indicating high growth opportunities, it could also signal overvalued firms. Higher growth opportunities tend to attract a wider cross-section of informationally-endowed traders, thereby leading to greater divergence in opinions and higher SIGOF. Short-sales constraints tend to drive out a fraction of the pessimistic traders from the overvalued stocks, thereby reducing divergence in opinions and lowering SIGOF. The net effect is an empirical issue.

Panel C of Table 3.4 stratifies the sample by ANAL and Size. SIGOF is low for small firms that are ignored by analysts and high for large firms that are followed by more analysts. The results suggest that even though size and ANAL are highly correlated, the univariate results observed in Table 3.3 continue to exist along both dimensions in the bivariate setup. Both larger firms, as well as firms followed by larger numbers of analysts, tend to display more volatile order-flow. Since larger firms are likely to attract a broader cross-section of investors, they are also likely to be the focus of more dispersed opinions concerning firm value.

The results so far consistently place SIGOF as a measure of divergence in opinions. The next set of results explores this association in more detail. Table 3.5 summarizes the Spearman correlation matrix of the key variables (Pearson correlations give similar results). The table provides the average monthly correlation coefficients, obtained by estimating cross-sectional correlations for every month from January 1993 to December 2003 and calculating their time-series averages.

The average pair-wise correlation between SIGOF and the per dollar adverse selection cost is -0.572. This lends some preliminary support to Hypothesis 1. The negative correlation between DVIA and SIGOF suggests that, at times of high order-flow variability, the market maker is less concerned about losses due to adverse selection. The average pair-wise correlation between SIGOF and trading volume (VOL) is 0.660. A positive correlation between trading volume and SIGOF implies that, on average, firms with more volatile order-flow will experience greater trading volume. The positive correlation between trading volume and SIGOF provides evidence in support of

Hypothesis 3. The correlation between variability in trading volume (σ_{Vol}) and order-flow variability (SIGOF) is 0.488. This lends preliminary support to Hypothesis 4.

Table 3.5 also provides some evidence in support of Hypotheses 5 and 6. The correlation between dispersion in analysts' forecasts and SIGOF is 0.036, which, though small, is statistically significant. This result suggests that firms with more volatile order-flow are also more likely to have more dispersed analysts' earnings forecasts. The positive and significant correlation between SIGOF and risk-adjusted returns (r) lends support to Hypothesis 6. This implies that periods of high order-flow variability are likely to be associated with higher returns.

To address the relation between SIGOF and future return, trading volume, spreads, and the adverse selection and inventory components of the spread, Table 3.6 presents the distribution of these variables across lagged SIGOF quintiles. At the end of each month, we sort the sample stocks into quintiles based on the level of SIGOF in the previous month. We compute the average $DVIA, DVINV, r, VOL$, quoted spread (QSPR) and proportional spread for each quintile in the following month. Panel A presents the time-series means of the monthly DVIA, r, DVINV, quoted and proportional spread, and VOL. Panel B presents the difference between the first and the fifth quintile portfolios and the difference in mean statistics.

We find a statistically significant negative relation between lagged order-flow variability and risk-adjusted stock returns (the fifth quintile portfolio return is 1.02% lower than the first quintile portfolio return). This result provides evidence in support of Hypothesis 7, suggesting that periods of high divergence in opinions are likely to be followed by declining prices and therefore lower returns. In the cross-section, firms which experience more volatile order-flow at time t-1 are likely to see a greater decline in time t returns.

We find a statistically significant negative relation between SIGOF and both adverse selection and inventory costs (i.e. a statistically significant difference between the fifth quintile portfolio mean and the first quintile portfolio mean). Kyle (1985) provides a useful framework for partially interpreting this result. A market-maker's adverse selection problem is directly proportional to the prior uncertainty about the firm's fundamentals and informed trader activities, and inversely proportional to uninformed trader activities. If more volatile order-flow can be interpreted as indicating more heterogeneous investors, it would also suggest greater uninformed trader activity, and thus, lower adverse selection. Given that higher SIGOF implies greater lack of synchronicity in buyer and seller trades, it is expected to increase the inventory management cost of the market-maker. The observed negative relation is puzzling.

Table 3.7 presents the results of the cross-sectional model, specified in Equation (3.4). We estimate the model by event month. Table 3.7 reports the time-series averages of the estimated slope coefficients and the t-statistics corresponding to the test $H_0 : \overline{\beta}_j = 0$.

The positive and significant coefficients on Vol and SigVol lend support to Hypotheses 3 and 4. The results suggest that periods of high volume and volume volatility are associated with high SIGOF. We find evidence for a positive association between dispersion in analyst forecasts and SIGOF. This provides support for Hypothesis

5. We also find a negative and significant association between the adverse selection cost of trading and SIGOF. This reinforces the support for Hypothesis 1. Once again, we find evidence of negative association between the inventory management cost and SIGOF and so reject Hypothesis 2. Overall, Table 3.7 provides strong support in favor of interpreting SIGOF as a measure of divergence in opinions.

The coefficient for number of analysts following the firm is found to be consistently positive. This suggests that firms with more analyst following are also firms with more volatile order-flow. This result can be interpreted in at least two ways. First, a larger number of analysts generates more firm-related data and, therefore, makes more information available for smart investors to trade on. This increases the gap between naïve and smart investors in the market, leading to more volatile order-flow. A second interpretation is that larger numbers of analysts are attracted to the markets with high order-flow variability because more demand exists for information in these markets.

Table 3.8 presents the results of the time-series model specified in Equation (3.5). We estimate the model for every stock in the sample. Table 3.8 reports the time-series averages of the estimated slope coefficients and the t-statistics corresponding to the test $H_0 : \overline{\beta}_j = 0$. We find a positive association between the proportional changes in risk-adjusted returns and the proportional changes in SIGOF. This result is consistent with the divergence in opinions interpretation of SIGOF. As the divergence in opinions increases, it drives out a fraction of the pessimists from the market, thereby inflating the stock price and leading to higher returns. This result provides evidence in support of Hypothesis 6.

We also find evidence of a positive contemporaneous association between changes in S&P 500 futures open interest and changes in the order-flow variability for the average stock (Hypothesis 8). Bessembinder, Chan, and Seguin (1996) suggest that open interest on the S&P 500 futures contract represents an empirical proxy for cross-sectional dispersion in traders' opinions about market information. The positive contemporaneous coefficient in Table 3.8 suggests that the average stock's SIGOF contains a market-wide component and points to the existence of commonality in SIGOF.

3.5.2 Commonality in order-flow variability

The above analysis hints at the existence of commonality in SIGOF. This section attempts to explore it further. First, we explore the nature of cross stock commonality in SIGOF. Then, we attempt to throw some light on its implications for co-movement in liquidity, adverse selection costs, and inventory costs.[21]

Following the methodology outlined in Section 3.6, Table 3.9 presents the statistics for the β_i coefficients from Equation (3.7). Approximately 83% of the individual β_i are positive, with 37% significant at the 5% one-tailed critical value. For the quoted spread, β_i is positive for approximately 93% of the stocks in the sample; 67% of these are statistically significant. The corresponding proportion of positive (positive and significant) β_i coefficients for proportional spreads, adverse selection costs and inventory costs are 96% (79%), 92% (58%) and 76% (27%), respectively. These results provide evidence of co-movement in SIGOF, liquidity, adverse selection costs, and inventory costs (consistent with Chordia, Roll, and Subrahmanyam, 2001). The average R^2 for the regressions are about 12% for the quoted spread, 16% for the proportional quoted spread, 6% for SIGOF, 6% for DVIA, and about 2% for DINV.

---

[21] Chordia, Roll, and Subrahmanyam (2001) present arguments supporting co-movement in inventory and adverse-selection costs. However, to the best of our knowledge no study has addressed their existence in a large sample.

Table 3.10 (Panel A) presents the results of estimating Equation (3.7), using the level of each variable instead of the proportional change. For SIGOF, 86% of the β coefficients are positive and 75% are positive and significant. In the case of the proportional spreads, we find that 92% of β are positive while 84.5% are positive and significant. Of the coefficients for the monthly quoted spreads, 95% are positive while 91% are positive and significant. Of the coefficients for the adverse selection cost component, 92% are positive, and 88% of the inventory cost component coefficients are also positive. These results are similar to the findings noted in Table 3.9 (based on changes). Table 3.10 provides additional evidence for the existence of systematic components of SIGOF contributing to co-movement in SIGOF.

A final issue that remains to be explored is the implications of the co-movement in SIGOF for co-movement in liquidity, adverse selection costs and inventory costs. Table 3.10 (Panel B) provides the results of estimating Equation (3.7) using the residuals from Equation (3.8). Comparing the statistics in Panel A with those in Panel B allows us to identify the contribution of SIGOF in QSPR, PQSPR, DVIA, and DVINV co-movement. The explanatory power of the regressions in Panel B is lower than those in Panel A. The adjusted R^2 declines from 63% to 27% for quoted spreads (QSPR), from 49% to 24% for proportional spreads (PQSPR), from 28% to 14% for DVIA, and from 34% to 16% for DVINV. These results suggest that commonality in liquidity and in trading cost is at least partially determined by order-flow variability or factors determining order-flow variability. These results are in favor of Hypothesis 9.

The results of the pair-wise correlation analysis exploring the contribution of SIGOF co-movement to co-movement in liquidity are presented in Table 3.11. The average (median) pair-wise correlation between quoted spreads is 0.5134 (0.6290). Controlling for the contemporaneous SIGOF (Equation 3.6), we find that the magnitude of the correlation drops to a mean (median) level of 0.2002 (0.2207). We find a similar decline in the pair-wise correlation for proportional spreads, where the mean (median) correlation drops from 0.4206 (0.4733) to 0.1901 (0.2158). The cross-stock mean (median) correlation in the adverse selection cost declines from 0.2640 (0.2502) to 0.0865 (0.0743), and the correlation in the inventory cost drops from 0.2945 (0.2731) to 0.1217 (0.0956). These results provide additional evidence to support Hypothesis 9, whereby order-flow volatility can partially explain co-movement in liquidity, adverse selection costs, and inventory costs.

Table 3.12 attempts to shed some light on the factors responsible for the commonality in order-flow variability. We run the co-movement analysis (as described in Equations 3.7 and 3.8) on SIGOF, using lagged volume volatility (SIGVOL), S&P 500 open interest (NOIC), risk-adjusted market returns, number of analysts (ANAL), market-to-book ratios, and dispersion in analysts' forecasts (DISP) as the Equation 3.8 control variables. The first column in Table 3.12 presents the co-movement in SIGOF (the base case). The remaining columns correspond to co-movement analysis using Equation 3.8 residuals, controlling for the various proxies for divergence in traders' opinions. As we control for the effects of various divergence in opinion proxies, we notice a decline in β_i from 0.854 to 0.228. The t-statistics declines from 35.5 to 7.03 after controlling for various proxies for divergence in opinions. The adjusted R^2 declines from 46.16% to 0.95%. The percentage of stocks with positive β_i also declines from 85.87% to 70.65%. These results suggest that divergence in opinions is, at least partially, responsible for the observed commonality in the order-flow variability and hence for the commonality in liquidity as well as in the adverse selection and the inventory costs of trading.

3.6 Conclusion

Chapter three explores the characteristics of order-flow variability and examines the relation between order-flow variability and various proxies for divergence in opinions among traders. Our results suggest that order-flow variability measures divergence in opinions among traders. This paper uses 'divergence in opinions' in a rather broad sense, in referring to the dispersed beliefs of traders in the market. This dispersion could result from either rational or irrational causes. Rational reasons include differences in information across traders and traders using different information updating functions. Divergence in opinion could arise due to the existence of irrational agents in the market.

The second part of this study attempts to explore the commonality in order-flow variability, liquidity, adverse selection costs, and inventory carrying costs. Our results suggest that order-flow variability, liquidity, adverse selection costs, and inventory carrying costs, tend to co-move across assets. We provide evidence suggesting that market-wide divergence in opinions among traders is responsible, at least partially, for this co-movement.

References

Admati, A.R., and P. Pfleiderer, 1988. A theory of intraday patterns: volume and price variability, *Review of Financial Studies*, 1, 3-40.

Amihud, Yakov, and Haim Mendelson, 1986. Asset pricing and the bid-ask spread, *Journal of Financial Economics* 17, 223–249.

Bessembinder, H., K. Chan and P.J. Seguin, 1996. An Empirical Examination of Information, Differences of Opinion, and Trading Activity, *Journal of Financial Economics*, Vol. 40, pp.105–34.

Brennan, Michael J., Tarun Chordia, and Avanidhar Subrahmanyam, 1998. Alternative factor specifications, security characteristics and the cross-sectional of expected returns, *Journal of Financial Economics,* 49, 345-373.

Choie, K., and S.J. Hwang, 1994. Profitability of Short-Selling and Exploitability of Short information, *Journal of Portfolio Management*, 20, 33-38.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2001. Market liquidity and trading activity, *Journal of Finance* 56, 501-30.

Easley, D. and M. O'Hara, 1992. Time and the process of security price adjustment, *Journal of Finance*, 47, 577-605.

Fama, E.F., and French, K.R., 1989. Business conditions and Expected returns on Stocks and Bonds, *Journal of Financial Economics,* 25, 23-49.

Fama, Eugene F., and James D. MacBeth, 1973. Risk, Return, and Equilibrium: Empirical Tests, *Journal of Political Economy*, 81, 607-636.

Gallant, A., P. Rossi, and G. Tauchen, 1992. Stock prices and volume, *Review of Financial Studies* 5, 199-242.

Glosten, L. R., and Harris, L. E., 1988. Estimating the components of the bid ask spread. *Journal of Financial Economics*, 21, 123-142.

Grossman, Sanford J., and Merton H. Miller, 1988. Liquidity and market structure, *Journal of Finance*, 43, 617-633.

Harford, Jarrad, and Kaul, Aditya, 2005. Correlated order-flow: Pervasiveness, sources, and pricing effects, *Journal of Financial and Quantitative Analysis,* 40, 29–55.

Hasbrouck, J. and D. Seppi, 2001. Common factors in prices, order flows and liquidity, *Journal of Financial Economics,* 59, 383-411.

Ho, T., and H. Stoll, 1983. The dynamics of dealer markets under competition, *Journal of Finance* 38, 1053-1074.

Kandel, E. and N. Pearson, 1995. "Differential interpretation of public signals and trade in speculative markets," *Journal of Political Economy*, 103, 831-872.

Karpoff, Jonathan M., 1987. The relation between price changes and trading volume: A survey, *Journal of Financial and Quantitative Analysis*, 22, 109–126.

Kyle, Albert, 1985. Continuous Auctions and Insider Trading, *Econometrica*, 53, 1315-36.

Lee, C.M.C., and Ready, M.J., 1991. Inferring trade direction from intraday data, *Journal of Finance*, 46, 733-746.

Lee, C.M.C., and Ready, M.J., 1991. Inferring trade direction from intraday data, *Journal of Finance,* 46, 733-746.

Lee, C.M.C., and Radhakrishna, B., 2000. Inferring investor behavior: Evidence from TORQ data, *Journal of Financial Markets,* 3, 83-111.

Lin, J., Sanger, and G., Booth, G., 1995. Trade size and components of the bid-ask spreads. *Review of Financial Studies* 8, 1153-1183.

Lo A., and J. Wang, 2000. Trading volume: definitions, data analysis, and implications of portfolio theory, *Review of Financial Studies* 13, 257-300.

Merton, Robert C., 1987. A Simple Model of Capital Market Equilibrium with Incomplete Information, *The Journal of Finance*, 42, 483-510

Milgrom, Paul, and N. Stokey, 1982. Information, Trade and Common Knowledge, *Journal of Economic Theory*, 26, 17-27

Miller, E.M., 1977. Risk, uncertainty and divergence of opinion, Journal of Finance 32, 1151–1168.

Neal, R., and Wheatley S., 1998. Adverse selection and bid-ask spreads: Evidence from closed-end funds, *Journal of financial markets,* 1, 121-149

Odders-White, Elizabeth R., 2000. On the occurrence and consequences of inaccurate trade classification, *Journal of Financial Markets*, 3, 205-332.

Pontiff, Jeffrey, 1996. Costly arbitrage: evidence from closed-end funds, *Quarterly Journal of Economics*, 111, 1135-1151.

Senchak, A.J., and L.T. Starks, 1993. Short-Sale Restrictions and Market Reaction to Short-Interest Announcements, *Journal of Financial and Quantitative Analysis*, 28, 177-194.

Appendix

Table 3.1: Distribution of firms across the sample period.

The count presents the number of firms in the given year for which all the data described in the table were available. All presented numbers are monthly averages. The numbers presented in parenthesis, are the standard deviations across the sample in the particular year. SIGOF is the standard deviation of the 15-minute order imbalance. IA is the adverse selection cost component of the spread and INV is the inventory cost component of the spread (Lin, Sanger and Booth, 1988). Vol is the monthly traded volume, *r* is the risk-adjusted monthly return, S is the market capitalization and ANAL is the average number of analysts providing earnings estimate.

Year	Count	SIGOF	IA	INV	VOL	r	MB	Size	ANAL	DISP
1993	1024	2.0449	0.3647	0.3588	38464.1231	0.0112	3.3509	2957022.82	13.3625	0.1452
		(1.283)	(0.195)	(0.222)	(60559.673)	(0.081)	(5.745)	(5710362.497)	(9.316)	(0.185)
1994	1176	2.0058	0.3954	0.3330	37890.9219	0.0023	2.9328	2682912.79	12.7174	0.1309
		(1.178)	(0.416)	(0.421)	(62307.669)	(0.075)	(4.747)	(5314560.587)	(9.083)	(0.257)
1995	1249	2.0850	0.3838	0.3360	41240.9121	0.0037	2.9418	2954618.12	12.1284	0.1332
		(1.351)	(0.27)	(0.28)	(68864.049)	(0.078)	(5.31)	(6189732.228)	(8.702)	(0.192)
1996	1321	2.2289	0.4000	0.3113	47536.4857	0.0081	3.0493	3431706.99	11.4832	0.1217
		(1.582)	(0.645)	(0.653)	(77325.688)	(0.081)	(4.832)	(7197557.044)	(8.235)	(0.169)
1997	1408	2.4401	0.4540	0.2361	56984.2854	0.0093	3.0276	4001075.08	10.6894	0.1113
		(1.818)	(0.462)	(0.47)	(99056.006)	(0.091)	(5.029)	(8747716.418)	(7.689)	(0.148)
1998	1398	2.6196	0.5069	0.1741	71378.6026	-0.0100	2.7896	4547333.39	10.2762	0.1088
		(2.041)	(0.561)	(0.565)	(120643.577)	(0.11)	(5.093)	(9912058.479)	(7.4)	(0.156)
1999	1359	2.8872	0.4835	0.1942	86380.2711	0.0008	2.4075	4907976.13	10.7654	0.1041
		(2.514)	(0.41)	(0.414)	(142119.863)	(0.122)	(4.102)	(10744081.698)	(7.476)	(0.155)
2000	1245	3.3258	0.4814	0.1866	112962.9103	0.0313	2.3686	5023442.19	10.6806	0.1041
		(3.076)	(0.775)	(0.779)	(179788.967)	(0.133)	(3.8)	(10459072.58)	(7.424)	(0.155)
2001	1162	4.6615	0.4593	0.1255	120635.3947	0.0215	2.2953	4755739.52	9.4567	0.1162
		(4.07)	(0.265)	(0.251)	(184400.014)	(0.11)	(3.928)	(9544059.283)	(6.808)	(0.176)
2002	1109	6.1435	0.4556	0.1217	125709.7064	0.0204	2.1486	4145072.55	8.1447	0.0888
		(5.118)	(0.609)	(0.603)	(188083.879)	(0.097)	(4.536)	(7838240.588)	(6.038)	(0.138)
2003	1081	7.3173	0.4247	0.1408	128425.5163	0.0167	2.0864	3624256.89	8.4979	0.1041
		(5.776)	(0.439)	(0.431)	(181431.229)	(0.087)	(5.644)	(6367909.02)	(6.336)	(0.199)

Table 3.2: Distribution of SIGOF by Industry.

Industry	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
Computer Manufacturing	3.794	3.705	4.335	4.560	4.910	4.762	6.329	8.050	9.628	11.044	10.941
Communication and electronic equipment	2.708	2.730	3.032	3.339	3.540	3.468	3.866	4.856	5.802	7.279	8.608
Oil and Gas	2.178	1.940	1.904	2.239	2.589	2.842	2.990	3.455	4.972	6.231	7.710
Financial institutions	2.418	2.369	2.577	2.792	3.266	3.616	3.820	4.653	7.217	9.237	10.149
Computer and Data Processing Services	2.567	2.786	2.910	3.747	3.674	3.438	3.785	3.959	7.703	9.916	11.234
Optical, Medical, and Scientific instruments	2.476	2.171	2.443	2.978	2.942	2.794	3.072	4.088	5.965	7.344	8.193
Retailers	2.457	2.208	2.369	2.635	2.651	3.072	3.382	3.983	5.996	8.074	8.841
Wholesalers	1.981	1.915	1.994	2.092	2.135	2.477	2.757	3.097	4.634	6.640	8.419
Miscellaneous manufacturing	2.394	2.331	2.466	2.605	2.899	3.129	3.596	4.099	5.776	7.941	8.891
Health care and HMOs	2.209	2.288	2.440	2.639	2.659	2.698	2.880	3.889	6.139	8.079	9.191
Drugs and Genetic engineering	3.292	2.996	3.175	3.929	4.524	4.567	5.100	6.452	9.007	10.696	12.484
Miscellaneous Services	2.120	2.148	2.213	2.481	2.399	2.683	3.178	3.589	4.767	6.828	8.399
Transportation and Public utilities	1.996	2.040	2.038	2.171	2.353	2.744	3.210	4.000	6.073	7.810	9.166
Mining	1.998	1.998	2.045	2.147	2.226	2.574	2.884	2.954	4.783	5.844	6.164
Construction	1.996	1.937	1.779	2.112	2.036	2.136	2.469	2.852	5.044	7.160	8.353
Others	2.011	1.935	1.994	2.142	2.509	2.969	3.449	4.168	5.898	7.611	8.931

Table 3.3: Time-series and cross-sectional variation in average SIGOF.

Panel A: Cross-section divided into firm size (Market capitalization) quintiles:

Size Quintiles	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
1	1.466	1.438	1.466	1.504	1.541	1.545	1.569	1.611	2.079	2.673	3.126
2	1.709	1.601	1.646	1.729	1.836	1.922	1.919	2.051	3.003	4.363	5.329
3	1.816	1.779	1.832	1.926	2.079	2.24	2.319	2.595	3.831	5.425	6.732
4	2.192	2.171	2.312	2.468	2.766	3.085	3.345	3.948	6.018	8.305	9.889
5	3.489	3.495	3.764	4.377	5.216	5.862	7.032	8.129	11.306	14.355	16.478

Panel B: Cross-section divided into firm market-to-book ratio quintiles:

MB Quintiles	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
1	1.991	1.874	1.911	1.96	2.074	2.161	2.289	2.423	3.181	4.46	5.449
2	1.949	1.911	1.954	2.013	2.205	2.318	2.393	2.718	4.151	5.835	7.022
3	2.191	2.1	2.147	2.352	2.586	2.795	3.105	3.553	5.344	7.198	8.68
4	2.391	2.307	2.444	2.654	2.976	3.303	3.624	4.24	6.387	8.62	10.102
5	2.966	2.88	3.067	3.426	3.868	4.342	5.198	6.174	8.418	10.802	12.543

Panel C: Cross-section divided into quintiles of number of analysts following the firm:

ANAL Quintiles	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
1	1.56	1.493	1.534	1.604	1.659	1.741	1.73	1.948	2.758	3.988	4.767
2	1.753	1.66	1.709	1.777	1.928	2.095	2.1	2.45	3.472	5.319	6.42
3	2.005	2.003	2.155	2.221	2.322	2.632	2.826	3.408	5.033	7.337	8.782
4	2.563	2.54	2.686	2.945	3.176	3.624	4.21	5.178	7.665	10.32	12.415
5	3.812	3.753	3.991	4.576	5.332	6.145	6.972	8.295	11.614	14.408	16.79

Table 3.4: Distribution of SIGOF.

Panel A: Distribution of SIGOF by market-to-book ratio and size quintiles:

		Quintiles of MB				
		1	2	3	4	5
Quintiles of size	1	1.846	1.883	1.969	2.110	2.068
	2	2.531	2.407	2.520	2.574	2.796
	3	3.083	3.238	3.224	2.958	3.549
	4	4.318	4.164	4.384	4.413	4.578
	5	7.166	6.748	6.529	7.227	7.252

Panel B: Distribution of SIGOF by market-to-book ratio and ANAL (number of analysts providing earnings forecasts for the firm) quintiles:

		Quintiles of MB				
		1	2	3	4	5
Quintiles of ANAL	1	2.038	2.187	2.255	2.357	2.549
	2	2.691	2.621	2.639	2.890	3.109
	3	3.191	3.255	3.380	3.363	4.225
	4	4.467	4.527	4.512	4.895	5.840
	5	5.339	5.759	6.380	7.268	7.469

Panel C: Distribution of SIGOF by size and ANAL quintiles:

		Quintiles of Size				
		1	2	3	4	5
Quintiles of ANAL	1	1.804	2.268	2.863	4.390	6.479
	2	2.105	2.559	3.020	3.938	6.244
	3	2.347	2.790	3.130	3.887	6.086
	4	2.596	3.750	3.737	4.461	6.600
	5	2.250	4.982	4.660	5.123	7.389

Table 3.5: **Non-Parametric Correlation Coefficients (Spearman's rank correlation).**

Time-series averages of the monthly cross-sectional correlation coefficients; SIGOF is the variability of 15-minute order-flow within a month; σ_{Vol} is the standard deviation of 15-minute trading volume, within a month. DVIA is the per dollar adverse selection cost. It is calculated as the adverse selection cost component times the effective spread divided by the trading price. Vol represents the total trading volume within the give month. Ret is raw holding period return while risk-adjusted return is calculated using the four factor model described in Equation 3.3, MB is the market-to-book ratio of the firm, ANAL is the total number of analysts providing earnings forecasts for a given firm while DISP is the dispersion in their forecasts. Size is the market capitalization of the firm.

	SIGOF	σ_{Vol}	DVIA	Vol	ret	ret (risk adjusted)	MB	ANAL	DISP	Size
SIGOF	1									
σ_{Vol}	0.488	1								
DVIA	-0.572	-0.376	1							
Vol	0.660	0.849	-0.614	1						
ret	0.027	0.024	-0.034	0.013	1					
ret (risk Adj.)	0.003	0.003	-0.020	-0.011	0.972	1				
MB	0.342	0.014	-0.335	0.269	0.102	0.091	1			
ANAL	0.713	0.247	-0.672	0.757	-0.002	-0.019	0.249	1		
DISP	0.036	0.011	-0.054	0.032	-0.016	-0.018	-0.177	0.066	1	
Size	0.782	0.396	-0.798	0.814	0.016	-0.004	0.386	0.784	-0.003	1

Table 3.6: **Stock return, trading volume, spread and components of spread across SIGOF quintiles.**

Distribution of the adverse selection cost per dollar traded (DVIA), Inventory cost per dollar traded (DVINV) risk-adjusted returns (r), and monthly trading volume (VOL) across **SIGOF** quintile portfolio. Each month stocks are sorted into five groups based on the level of **SIGOF** for the previous month. We compute average DVIA, DVINV, quoted and proportional spreads, risk-adjusted return and trading volume for each quintile portfolio. Fama French four-factor model is used for calculating the risk-adjusted returns. Panel A reports the time-series distribution of DVIA, DVINV, quoted and proportional spreads, risk-adjusted returns, trading volume. Panel B reports the comparison of equality of means for portfolio 1 (lowest level of SIGOF) and portfolio 5 (highest level of SIGOF).

Panel A:

Quintiles of Sigma OF		Risk adj ret (VW)	Vol (,000)	DVIA	DVINV	QSPR	PQSPR
1	Mean	0.0113	843,622	0.0054	0.0035	0.1641	0.0122
	Median	0.0127	410,742	0.0057	0.0030	0.1698	0.0130
	Std Deviation	0.0272	886,062	0.0015	0.0020	0.0402	0.0028
2	Mean	0.0108	661,387	0.0050	0.0027	0.1682	0.0107
	Median	0.0110	598,829	0.0053	0.0021	0.1831	0.0119
	Std Deviation	0.0261	198,318	0.0015	0.0017	0.0460	0.0030
3	Mean	0.0088	1,574,477	0.0038	0.0020	0.1561	0.0084
	Median	0.0069	1,378,140	0.0042	0.0014	0.1725	0.0094
	Std Deviation	0.0272	634,582	0.0014	0.0014	0.0577	0.0031
4	Mean	0.0059	4,380,680	0.0026	0.0015	0.1429	0.0062
	Median	0.0059	3,652,972	0.0030	0.0011	0.1541	0.0069
	Std Deviation	0.0295	2,345,468	0.0010	0.0010	0.0602	0.0027
5	Mean	0.0011	21,191,553	0.0014	0.0011	0.1300	0.0040
	Median	0.0013	16,706,801	0.0016	0.0008	0.1398	0.0040
	Std Deviation	0.0234	12,669,757	0.0005	0.0008	0.0570	0.0019

Panel B: Test for difference between the 1st and the 5th quintile:

	-0.0102	20,347,931	-0.0040	-0.0024	-0.0341	-0.0082
	-3.2266	18	-28.9846	-12.9868	-5.5806	-27.3132

Table 3.7: Attributing SIGOF to firm and trading characteristics.

$$SIGOF_{i,t} = \alpha_t + \beta_{1,t} \times \ln\left(\sigma_{VOL,i,t}\right) + \beta_{2,t} \times \ln\left(Vol_{i,t}\right) + \beta_{3,t} \times \ln\left(Size_{i,t}\right) + \beta_{4,t} \times IA_{i,t} + \\ + \beta_{5,t} \times INV_{i,t} + \beta_{6,t} \times \ln\left(Anal_{i,t}\right) + \beta_{7,t} \times DISP_{i,t} + \beta_{8,t} \times MB_{i,t} + \beta_{9,t} \times \ln(ret_{i,t}) + \varepsilon_{i,t}$$

The table reports the time-series averages of the slope coefficients. The t-statistics corresponds to the test $Average(\beta) = 0$. The table also reports the cross sectional mean of the adjusted R^2 from the time-series regressions. The explanatory variables are, risk adjusted return, dispersion in analysts' forecasts (DISP) , market to book ratio of the firm (MB) and natural logarithm of: trading volume (Vol), 15-minute variability in trading volume within the month (σ_{Vol}), and number of analysts' providing earnings forecasts (ANAL). IA and INV are the adverse selection and inventory cost component of the spread respectively. The constant term is not reported.

ln(σ_{Vol})	ln(Vol)	ln(Size)	IA	INV	ln(Anal)	DISP	MB	Risk Adj Ret	Mean Adj R^2
1.358									0.476
(20.47)									
	1.131								0.564
	(20.32)								
		1.275							0.561
		(20.62)							
			[illegible]						0.042
			(-3.87)						
				-0.905					0.016
				(-4.78)					
0.598	0.546	0.699							0.637
(21.64)	(12.36)	(26.58)							
		1.281			0.745	0.396	-0.014		0.614
		(23.86)			(11.05)	(13.01)	(-1.70)		
0.15	0.783	0.629	-1.575		0.181	0.46	-0.034	-0.16	0.677
(10.44)	(14.84)	(27.7)	(-7.65)		(4.14)	(19.16)	(-4.19)	(-1.63)	
0.148	0.831	0.619		-0.509	0.183	0.419	-0.025	-0.155	0.673
(10.2)	(15.14)	(27.63)		(-2.32)	(4.21)	(17.39)	(-3.2)	(-1.64)	
0.146	0.792	0.619	-2.343	-1.381	0.177	0.437	-0.027	-0.188	0.679
(10.05)	(14.53)	(28.12)	(-10.55)	(-7.14)	(4.19)	(18.08)	(-3.42)	(-1.99)	

*t-statistics are given in the parenthesis bellow the coefficients.

Table 3.8: Attributing time-series changes in SIGOF to changes in systematic and firm specific factors.

$$DSIGOF_{i,t} = \alpha_i + \beta_{1,i} \times DNOIC_{i,t} + \beta_{2,i} \times DVol_{i,t} + \beta_{3,i} \times DSigVol_{i,t} + \beta_{4,i} \times r_{i,t} + \beta_{5,i} \times DANAL_{i,t} + \beta_{6,i} \times DDISP_{i,t} + \beta_{7,i} \times DMB_{i,t} + \varepsilon_{i,t}$$

Monthly proportional change in the individual stock's order-flow variability (SIGOF) is regressed in time-series on the cotemporaneous explanatory variables. The table reports the cross sectional averages of the slope coefficients. The t-statistics corresponds to the test $Average(\beta) = 0$. The table also reports the cross sectional mean of the adjusted R^2 from the time-series regressions. The explanatory variables are, risk adjusted return and monthly proportional change in: number of S&P open interest contracts (NOIC), trading volume (Vol), monthly volatility of the trading volume (SigVol), the number of analysts' providing earnings forecasts (ANAL), dispersion in analysts' forecasts (DISP), and market-to-book ratio of the firm (MB).

DNOIC	DVol	DSigVol	beta adj ret (VW)	DANAL	DDISP	DMB	Mean Adj R2
0.005							0.018
(1.88)							
	0.157						0.304
	(63.95)						
		0.062					0.224
		(47.46)					
			0.288				0.038
			(17.05)				
				-0.012			0.003
				(-0.89)			
					0.009		0.009
					(2.13)		
		0.064	0.239		-0.019	0.02	0.254
		(23.63)	(2.05)		(-0.78)	(0.18)	
		0.061	0.323	-0.042	-0.034	-0.06	0.257
		(19.33)	(4.26)	(-0.97)	(-0.81)	(-0.77)	
0.023		0.061	0.258	-0.045	-0.037	-0.051	0.258
(0.55)		(18.54)	(4.19)	(-1.04)	(-0.88)	(-0.71)	

***t-statistics are given in the parenthesis bellow the coefficients.**

Table 3.9: Market-wide commonality in SIGOF and liquidity.

Monthly proportional change in individual stock's order-flow variability (SIGOF) is regressed in time-series on proportional change in the equal-weighted average order-flow variability for all stocks in the sample (the 'market').

$$DSIGOF_{i,t} = \alpha_i + \beta_i \times DSIGOF_{M,t} + \gamma_{i,1} r_{m,t-1} + \gamma_{i,2} r_{m,t} + \gamma_{i,3} r_{m,t+1} + \gamma_{i,4} \ln\left(\frac{P_{i,t,\max}}{P_{i,t,\min}}\right) + \varepsilon_{i,t}$$

The right hand side control variables include a lead and a lag market return $\left(r_{m,t+1} \text{ and } r_{m,t-1}\right)$, and a measure of monthly volatility (Natural logarithm of the ratio of the maximum stock price to the minimum stock price in the given month). The procedure is repeated for two liquidity measures: QSPR (the quoted spread) and PQSPR (the proportional quoted spread), proportional change in monthly adverse selection cost of trading (DDVIA) and the monthly proportional change in inventory cost incurred by the market maker (DDVINV).

The letter D denotes proportional change. Therefore for measure M, $DM_t = \left(M_t - M_{t-1}\right)/M_{t-1}$

Cross-sectional averages of time-series slope coefficients and the corresponding t-statistics are reported. '%Positive' reports the percentage of positive slope coefficients, while '% + Sig' gives the percentage with t-statistics greater than +1.29 (10% critical level for one tail test)

	DQSPR	DPQSPR	DSIGOF	DDVIA	DDVINV
Adj R2 Mean	11.80%	15.94%	5.74%	6.32%	1.77%
Adj R2 Median	7.89%	13.70%	3.88%	3.81%	0.11%
% Positive	92.46	96.01	82.72	92.09	76.30
% + Sig.	67.30	78.28	36.72	58.33	26.59
% Negative	7.54	3.99	17.28	7.91	23.70
% - Sig.	0.53	0.13	0.90	0.40	2.54

Table 3.10: Market-wide commonality in levels of liquidity.

(as measured by Quoted spread (QSPR) and proportional quoted spread (PQSPR)), adverse selection cost per dollar of trade (DVIA) and Inventory cost per dollar of trade (DVINV):
Monthly levels of individual stock's order-flow variability (SIGOF) is regressed in time-series on the equal-weighted average order-flow variability for all stocks in the sample (the 'market').

$$SIGOF_{i,t} = \alpha_i + \beta_i \times SIGOF_{M,t} + \gamma_{i,1} r_{m,t-1} + \gamma_{i,2} r_{m,t} + \gamma_{i,3} r_{m,t+1} + \gamma_{i,4} \ln\left(\frac{P_{i,t,\max}}{P_{i,t,\min}}\right) + \varepsilon_{i,t}$$

The right hand side control variables include a lead and a lag market return $\left(r_{m,t+1} \text{ and } r_{m,t-1}\right)$, and a measure of monthly volatility (Natural logarithm of the ratio of the maximum stock price to the minimum stock price in the given month). The procedure is repeated for two liquidity measures: QSPR (the quoted spread) and PQSPR (the proportional quoted spread), monthly adverse selection cost of trading (DVIA) and the monthly inventory cost incurred by the market maker (DVINV).

Panel A reports the cross-sectional averages of time-series slope coefficients beta(β_i), and the corresponding t-statistics. '%Positive' reports the percentage of positive slope coefficients, while '% + Sig' gives the percentage with t-statistics greater than +1.29 (10% critical level for one tail test)

Panel B repeats the analysis using the residuals from:

$$M_{i,t} = \alpha_{1,i} + \alpha_{2,i} SIGOF_{i,t} + \varepsilon_{i,t}$$

Where $M_{i,t}$ is a general representation for the quoted spread (QSPR), proportional quoted spread (PQSPR), adverse selection cost per dollar traded (DVIA) and the inventory cost per dollar traded (DVINV), for firm i in month t.

Panel A: Commonality in liquidity, adverse selection cost and Inventory cost

	QSPR	PQSPR	SIGOF	DVIA	DVINV
Adj R2 Mean	63.02%	49.27%	46.16%	31.91%	34.06%
Adj R2 Median	74.08%	56.37%	49.26%	27.81%	30.24%
% Positive	95.19	91.77	85.88	91.72	87.91
% + Sig.	90.51	84.48	74.87	79.82	74.32

Panel B: Liquidity comovement, controlling for SIGOF

	QSPR	PQSPR	SIGOF	DVIA	DVINV
Adj R2 Mean	27.47%	24.19%		14.36%	15.63%
Adj R2 Median	26.64%	23.46%		11.91%	12.14%
% Positive	91.219	91.113		91.695	83.602
% + Sig.	81.62	78.87		75.17	63.45

Table 3.11: The contribution of co-movement in SIGOF to co-movement in liquidity.

Panel A presents the mean and the median pair-wise correlation, run across all 3870 firms in the sample. All pairs with less than 20 observations are omitted from the analysis.

Panel B presents the pair-wise correlation between the same set of variables, controlling for the effect of SIGOF. We regress the average monthly liquidity measures on contemporaneous SIGOF for the stock and examine the cross stock correlation of the residuals $(\varepsilon_{i,t})$ from the following model: $M_{i,t} = \alpha_{1,i} + \alpha_{2,i} SIGOF_{i,t} + \varepsilon_{i,t}$

Where $M_{i,t}$ is a general representation for the monthly average quoted spread (QSPR), proportional quoted spread (PQSPR), adverse selection cost per dollar traded (DVIA) and the inventory cost per dollar traded (DVINV), for firm i in month t.
This analysis helps to identify the contribution of co-movement in SIGOF to co-movement I liquidity.

Panel A

	QSPR	PQSPR	SIGOF	DVIA	DVINV
Mean Correlation	0.5134	0.4206	0.4362	0.2640	0.2945
Median Correlation	0.6290	0.4733	0.4587	0.2502	0.2731
Number of Observations	6,390,810	6,390,810	6,390,810	6,390,810	6,390,810

Panel B

	QSPR	PQSPR	SIGOF	DVIA	DVINV
Mean Correlation	0.2002	0.1901		0.0865	0.1217
Median Correlation	0.2207	0.2158		0.0743	0.0956

Table 3.12: Some explanations for market wide commonality in SIGOF.

Monthly individual stock's order-flow variability (SIGOF) is regressed in time-series on a set of lagged control variables. The set of control variables are: SIGOF, volume volatility, number of S&P open interest contracts (NOIC), value weighted market return, market capitalization of the firm (Size), trading volume (Vol), number of analysts providing earnings' forecasts for the firm (ANAL), Market to Book ratio (MB), and dispersion in analysts' forecasts (DISP).

We use the residuals $(\varepsilon_{i,t})$ from

$$SIGOF_{i,t} = \alpha_{1,i} + \sum_j \alpha_{2,i,t} \left(Control\ Variable_j \right)_{i,t} + \varepsilon_{i,t}$$

to run co-movement analysis, using the equation:

$$\varepsilon_{i,t} = \alpha_i + \beta_i \times \varepsilon_{M,t} + \gamma_{i,1} r_{m,t+1} + \gamma_{i,2} r_{m,t-1} + \gamma_{i,3} \ln\left(\frac{P_{i,t,\max}}{P_{i,t,\min}} \right) + \psi_{i,t}$$

The table presents, the cross-sectional averages of time-series slope coefficients beta (β_i), and the corresponding t-statistics. '%Positive' reports the percentage of positive slope coefficients, while '% + Sig' gives the percentage with t-statistics greater than +1.29 (10% critical level for one tail test)

	No Controls	SIGOF and SIGVOL	SIGOF, ln(NOIC), market ret.	SIGOF, ln(Size), Ln(Vol), ln(ANAL), MB, DISP, SIGVOL	SIGOF, ln(Size), Ln(Vol), ln(ANAL), MB, DISP, SIGVOL, Ln(NOIC), market ret.
	SIGOF	SIGOF	SIGOF	SIGOF	SIGOF
Mean beta	0.854	0.895	0.773	0.633	0.228
t-stat	35.506	37.212	37.591	26.550	7.034
Adj R2 Mean	46.16%	6.72%	6.96%	2.78%	0.95%
Adj R2 Median	49.26%	4.09%	3.84%	1.84%	-0.55%
% Positive	85.877	84.864	85.670	72.025	70.650
% + Sig.	74.87	57.44	55.80	19.79	16.54

Chapter 4

Firm opacity and financial market information asymmetry

4.1 Introduction

Information asymmetry between the firm and the market affects the firm's ability to raise external capital (Myers and Majluf, 1985).[22] Information asymmetry between investors affects liquidity (Kyle 1985; Admati and Pfleiderer, 1988) and hence, could also affect the availability of capital to the firm (Amihud and Mendelson, 1988). Nevertheless, few attempts have been made to relate these two types of information asymmetry. This study investigates how information asymmetry between a firm and an investor (hereafter called firm-to-investor IA) is related to the information asymmetry among investors (hereafter called inter-investor IA).

A firm needs to understand the relation between the firm-to-investor IA and the inter-investor IA as both will affect the firm's investment abilities. Although low levels of both information asymmetries might be desirable, it is not always possible for a firm to achieve this. A firm can reduce the level of its firm-to-investor IA through its transparency and disclosure decisions (though this can be expensive). Inter-investor IA, however, depends on such factors as the uncertainty about the value of the underlying asset and the trading activity of uninformed traders (Kyle, 1985), as well as the potential for some traders to process firm disclosures into superior information (Lundholm, 1991; Kim and Verrecchia, 1994). While a firm's decision to reduce firm-to-investor IA could reduce the uncertainty about the value of the underlying asset, its impact on the other two factors is not clear.

---

[22] Myers and Majluf (1984) suggested that the existence of firm-to-investor information asymmetry implies that managers (firm insiders) possess superior knowledge (relative to outsiders) as to future positive NPV investment opportunities. In such a situation, the market will assess such companies as having zero (or low) growth opportunities.

Some research (Diamond, 1985 and Hakansson, 1977) suggests that reduction in firm-to-investor information asymmetry leads to a reduction in the expected net benefit to investors with private information, thereby reducing their incentives to find the information in the first place. Thus, lower firm-to-investor information asymmetry should imply a lower risk of trading with an informationally-endowed trader. This reduces the level of inter-investor information asymmetry. Nevertheless, other studies (Lundholm, 1991; Kim and Verrecchia, 1994; Kandel and Pearson, 1995) suggest that more and better quality information releases by a firm provide more material to those investors looking to process public signals to create private benefits. If this is the case, lowering firm-to-investor information asymmetry may result in higher inter-investor information asymmetry. Huson and MacKinnon (2003) provide evidence in support of the latter group of studies. They show that when spin-offs increase a firm's focus, they lead to an increase in the informational gap between informed and uninformed investors. Similarly, Krinsky and Lee (1996) find that the adverse selection cost component of the bid-ask spread increases around earnings announcements.

Given the alternate views, it is not clear whether, on average, a reduction in firm-to-investor information asymmetry will lead to a reduction or to an increase in the adverse selection cost faced by the uninformed investors (inter-investor information asymmetry). This study finds evidence for a curvilinear relation between the adverse selection cost of trading and the firm-to-investor information asymmetry. As the level of firm-to-investor information asymmetry increases, the adverse selection cost of trading rises until a certain point, when it starts to decline. This result is intuitively appealing. If a firm is completely transparent, all market participants know everything about the firm and hence, the adverse selection should be zero.[23] If the firm is completely opaque, all participants are uninformed and hence the adverse selection problem should again become zero. Somewhere between the two extremes, the adverse selection cost attains a maximum.[24]

This study points towards various possible effects of a firm's transparency and disclosure-related decisions. A marginal increase in the level of transparency of a firm with a very high firm-to-investor IA could lead to an increase in its inter-investor IA. This might lead to reduced liquidity and possibly to a higher cost of capital. Similarly, a firm with a very high inter-investor IA might be able to increase the liquidity of its stocks by either reducing or increasing its firm-to-investor IA. Our results also suggest that increased firm transparency need not always be advantageous to the average investor in the market. Thus, this study also adds to the literature concerning the impact of market regulations pertaining to disclosure quality and firm transparency.

The remainder of chapter four is organized as follows. In Section 4.2, we develop the hypothesis and briefly discuss the research methodology. Section 4.3 describes the adverse selection cost of trading measure that we have chosen. In Section 4.4, we review the different proxies for firm-to-investor information asymmetry and define the various explanatory variables used in this study. Section 4.5 describes the data and sample used in this study, while Section 4.6 discusses the empirical findings. Section 4.7 attempts to throw some new light on the effect of focus enhancing corporate spin-offs on inter-investor information asymmetry. Section 4.8 concludes.

---

[23] A transparent firm is one with low to nil firm-to-investor information asymmetry. As the level of firm-to-investor information asymmetry declines, firms will become less transparent (or more opaque). This paper uses transparency and firm-to-investor information asymmetry synonymously. Opacity is the antonym of transparency.

[24] A caveat is in order here. For very opaque firms, the point of equilibrium for inter-investor IA will be determined through the interplay of search costs and the economic value of information. Therefore, the level of inter-investor information asymmetry might be non-zero.

4.2 Research Methods

The market microstructure literature views bid-ask spreads as the sum of three different costs incurred by the market-maker: inventory cost, order processing cost, and adverse selection cost. The inventory cost arises because the market-maker is forced to hold a non-diversified portfolio, which exposes the market-maker to non-systematic risks (Demsetz, 1968; Ho and Stoll, 1981). The market-maker incurs the order-processing cost in the process of making the market for a given security. The adverse selection cost results from the information asymmetry between informed traders in the market and the uninformed market-maker.

Information asymmetry in the stock market occurs when one or more investors either possess private information (Kyle, 1985) or are better able to process public information about the firm (Kim and Verrecchia, 1994, 1997). When a market-maker trades with the informed investors, the market-maker will lose money. He protects himself from the losses by building a non-zero adverse selection cost component (λ) into the bid-ask spread. Existing research shows that market-makers widen their bid–ask spreads when they suspect a high level of information asymmetry (Copeland and Galai, 1983; Glosten and Milgrom, 198; Venkatesh and Chiang, 1986). Since λ is the result of information asymmetry between the informed traders and the uninformed market-maker, a measure of information asymmetry is usually interpreted to exist between traders in the financial market (inter-investor). However, put in terms of the true value of the stock, the magnitude of λ will not only be a function of the information asymmetry surrounding the stock's true value, but also of the probability that the informed traders can capitalize on that asymmetric information.

The fact that adverse selection cost depends on the ability of the informed trader to capitalize on the asymmetry surrounding the true value of the stock has implications for the association between the transparency of a firm and the adverse selection cost of trading in the financial market. If a firm is completely transparent, no uninformed traders will be present and, by definition, the adverse selection cost of trading should be zero. For an absolutely opaque firm, in contrast, the absence of any informed traders should drive the adverse selection cost to zero. Although the locus of the association between the two extremes remains unexplored, it should contain the point of the maximum.

A caveat is in order here, as the above argument draws upon extreme cases of transparency and opacity. In the equilibrium state, the adverse selection problem is a function of the trading activity of informed and uninformed traders. The trading activity of the informed traders is determined through the interplay of their search costs and the economic value of their information.[25] As a firm becomes increasingly opaque, the search costs will increase. Theoretically, thee costs are infinite for absolutely opaque firms and zero for absolutely transparent firms. The economic value of the information is limited for at least two reasons. First, the price impact of trades limits potential gains. Second, capital constraints could also limit an investor's ability to take advantage of positive news and risky short sales could limit the potential gains from negative news. Thus, limited potential gains and increasing search costs should lead to a decline in the informed traders' activity and, therefore, to a decline in the adverse selection problem.

---

[25] Search cost refers to the cost incurred by an investor to find economically significant information. Economic value refers to the investor's potential profits from trading using this information.

The argument above suggests a curvilinear association between firm-to-investor IA and inter-investor IA. While the logic of this argument does not imply a functional form of the association, it does provide two characteristics to guide us in that direction. First, the two extreme cases (absolutely transparent and absolutely opaque) should constitute the minima, and, second, the maxima should lie between these two extremes.

4.2.1 Box-Cox transformations

On occasions, to empirically determine the correct functional form which is not specified, the family of power transformations, introduced by Box and Cox (1964), is used. It as and given by:

$$y^{(\theta)} = \begin{cases} (y^{\theta}-1)/\theta & \theta \neq 0 \\ \ln(y) & \theta = 0 \end{cases} \tag{4.1}$$

The transformed variables can be included in a linear function to specify and estimate a generalized model of the form:

$$y^{(\theta_1)} = \beta_1 + \beta_2 X_2^{(\theta_2)} + \ldots + \beta_k X_k^{(\theta_k)} + \varepsilon \tag{4.2}$$

Assuming ε_t to be normally distributed with mean zero and variance σ^2, Equation (4.2) is estimated by maximizing the non-linear likelihood function given by:

$$L\left(\beta_t, \theta_t, \sigma^2; X, y\right) = \prod_{t=1}^{T} f\left(\varepsilon_t\right) y_t^{\theta_t - 1} \tag{4.3}$$

$$\text{where } f\left(\varepsilon_t\right) = \left(2\pi\sigma^2\right)^{\frac{-1}{2}} \exp\left({}^{-\varepsilon_t^2}\!\big/_{2\sigma^2} \right)$$

We expect the estimated functional form to be characterized by the two properties laid out in the previous section.

4.2.2 *Polynomial regression*

The Box-Cox transformations are rather restrictive by construction. They allow for $\theta \leq 0$ if and only if the variable being transformed is non-negative at all points. Furthermore, the practice of estimating θ, and then performing inference about β, as if θ were known, has been criticized in the econometrics and statistics literature (Amemiya and Powel, 1981; Bickel and Doksum, 1981). Since the central issue being addressed in this paper does not strictly require the estimation of the functional form, we propose an alternative approach to test for the existence of the hypothesized relation between firm-to-investor and inter-investor information asymmetries. The method of choice is a simple polynomial regression. Based on the two guidelines, we propose a simple quadratic functional form:[26]

$$\lambda = \alpha + \beta_1 \times IA + \beta_2 \times IA^2 + \gamma(\mathit{other\ controls}) + \varepsilon \tag{4.4}$$

---

[26] We do not include additional terms such as square root, cubic, quartic, etc. in model 4 because this exercise is simply to demonstrate a curvilinear relationship and to that end, we only require the existence of the second order derivative for the specified function. A simple quadratic polynomial is therefore sufficient.

Here, λ is the adverse selection cost of trading in the firm's security and IA is the measure of information asymmetry between the firm and the market. Assuming that IA is either increasing or decreasing monotonically with opacity of the firm, we expect $\beta_1 \geq 0$ and $\beta_1 \leq 0$.

4.3 Measure of Adverse Selection Cost of Trading

This paper uses the adverse selection cost component of spread, as described by the Glosten and Harris (1988) model, to measure information asymmetry between investors in the stock market.[27] In the Glosten and Harris (1988) model, the adverse-selection, the inventory-holding and order-processing components, are expressed as a linear function of transaction volume. The model is described as follows:

$$\Delta P_t = c_0 \Delta I_t + c_1 \Delta I_t V_t + z_0 I_t + z_1 I_t V_t + \varepsilon_t \tag{4.5}$$

In this case, I_t is a trade indicator that equals 1 if the transaction is buyer-initiated, and -1 if it is seller-initiated; P_t is the transaction price at time t; V_t is the volume traded at time t; and ε_t captures public information innovations and errors. In this model, the adverse-selection component is $2(z_0 + z_1 V_t)$, and other components (inventory-holding and order-processing components) are measured as $2(c_0 + c_1 V_t)$. We use the average transaction volume for the stock to obtain an estimate of the adverse-selection component as a percentage of the bid-ask spread:

---

[27] We have also run the analysis using the adverse selection cost component as proposed by Lin, Sanger and Booth (1995) and Neal and Wheatley (1998). Our results are robust to the methodology selected. For the sake of brevity, I report only the results corresponding to λ, as computed using the Glosten and Harris (1998) algorithm.

$$\lambda = \frac{2\left(z_0 + z_1\bar{V}\right)}{2\left(c_0 + c_1\bar{V}\right) + 2\left(z_0 + z_1\bar{V}\right)} \times 100 \qquad (4.6)$$

We follow the Lee and Ready (1991) procedure for classifying trades. According to this algorithm, a trade is classified as buyer- (seller-)initiated if the transaction price is closer to the ask (bid) price of the prevailing quote. The quote must be at least five seconds old. If the trade is exactly at the midpoint of the quote, a "tick test" classifies the trade as buyer- (seller-)initiated if the last price change prior to the trade is positive (negative). Since the trade direction is inferred from the available information and not observed, some assignment error is inevitable; hence, the resulting order-flow data are estimates. Nevertheless, as shown by Lee and Radhakrishna (2000) and Odders-White (2000), the Lee and Ready (1991) algorithm is largely accurate; thus, inferences based on the estimated order-flow should be reliable.

4.4 Proxies for Firm-to-Investor Information Asymmetry

By its nature, the magnitude of firm-to-investor information asymmetry cannot be directly observed. In the existing literature, various proxies have been devised for measuring this information asymmetry. The measures used in this study may be broadly classified into three categories. The first set of measures is based on the quality of the firm's disclosures; the second category of proxies is based on firm characteristics; and the third category consists of proxies based on the precision of analysts' earnings forecasts. While the proxies based on firm characteristics are direct measures, the analyst-based proxies are indirect measures. To the extent that analysts rely on their understanding of the firm to generate earnings forecasts, we propose that the above hypothesis should hold true not only for the direct measures but also for the analyst-based indirect measures of the firm-to-investor information asymmetry.

This section discusses the proxies used in this study and their implications for the adverse selection cost of trading (inter-investor information asymmetry). We discuss the control variables in Section 4.4.4.

4.4.1 Proxies based on disclosure quality

A firm's choice of disclosure quality determines the distribution of firm-specific information among investors. Firms with higher disclosure quality are more likely to release timely forward-looking information. Therefore, these firms are likely to be more transparent than corresponding firms with lower disclosure quality. We use two measures of disclosure quality: the association of management and research (AIMR) disclosure scores, and S&P transparency and disclosure (T&D) scores for financial transparency and disclosure.

4.4.1.1 AIMR scores

The Association for Investment and Management Research (AIMR) has published the annual rankings of corporate disclosure practices for all years between 1982 and 1996. These scores have been widely used in academic research as an empirical proxy for disclosure quality.[28] According to the AIMR, these scores measure a firm's effectiveness in communicating with investors, and the extent to which its aggregate disclosure ensures that investors have the information necessary to make an informed judgment. According to Bushee and Noe (2000), one problem in using the AIMR database is that different industries are rated on different scales, since the analysts within each industry are only responsible for that industry's rankings. Therefore, scores across industries are not comparable. In addition, raw disclosure scores across time are not comparable. To address this problem we follow Bushee and Noe (2000), and convert raw disclosure scores into percentile ranks within each industry-year.[29]

---

[28] Other studies that have used the AIMR disclosure scores include Botosan and Plumlee (2002); Gelb and Zarowin (2002); Bushee and Noe (2000); Healy, Hutton, and Palepu (1999); Lang and Lundholm (1996); and Lundholm and Myers (2002). A detailed description of the data can be found in Bushee and Noe (2000).

4.4.1.2 The S&P T&D scores

The S&P Transparency and Disclosure (T&D) scores are obtained from Standard & Poor's Transparency and Disclosure dataset (published October 16, 2002). The scores are developed by collecting data from the annual reports (financial year ending 2002), 10-Ks, and proxy filings of 460 of the S&P 500 companies, based on 98 possible attributes, broadly classified into three major categories: (1) Ownership structure and investor rights, (2) Financial transparency and information disclosure, and (3) Board and management structure and process. Higher scores reflect the fact that a greater number of the attributes are present in the firm's disclosure.[30] S&P is careful to note that their rankings assess only the existence of a particular disclosure item. Although they rule out any attempt to assess the quality of the disclosure, Patel and Dallas (2002) document significant correlations between T&D rankings of US firms and determinants of expected returns such as market risk, size, and price-to-book ratio. These scores have been used in the existing literature as an alternative disclosure ranking metric for the discontinued AIMR rankings of disclosure, discussed in the previous sub-section.[31]

---

[29] Other research using this method include Lang and Lundholm (1993; 1996); and Healy, Hutton, and Palepu (1999).

[30] The T&D study focused on several issues such as which companies were providing the most extensive disclosure in their basic corporate filings, and which companies had disclosed above and beyond what the law requires. See Patel and Dallas (2002) for a detailed description.

4.4.2 Proxies based on firm characteristics

4.4.2.1 Discretionary accruals

Discretionary accruals have been widely used in tests of earnings management and market efficiency (Defond and Jiambalvo, 1994; Rees, Gill and Gore, 1996; Teoh, Welch, and Wong, 1998). Earnings management studies "examine whether managers act as if they believe users of financial reporting data can be misled into interpreting reported accounting earnings as equivalent to economic profitability." (Fields, Lys, and Vincent, 2001, p. 279). In other words, the managers' earnings management ability is directly related to their ability to confuse investors. Therefore, *ceteris paribus,* firms with higher potential for earnings management should be more opaque, and vice versa.

We estimate discretionary accruals using the cross-sectional version of the Jones (1991) model, as in Defond and Jiambalvo (1994):[32]

$$\frac{TotAcc_t}{TA_{t-1}} = \beta_1 \times \frac{1}{TA_{t-1}} + \beta_2 \times \frac{\Delta REV_t}{TA_{t-1}} + \beta_3 \times \frac{GPPE_t}{TA_{t-1}} + \varepsilon_t \qquad (4.7)$$

where $TotAcc_t$ is total accrual in year t; ΔREV_t is the difference between the revenue (data # 12) in year t and year t-1; $GPPE_t$ is gross property plant and equipment (data # 7), at the end of year t; and TA_{t-1} is total assets (data # 6) at the end of year t-1. The residual ε_t represents the discretionary portion of the total accruals at $TotAcc_t$. Following Balsam, Chen, and Sankaraguruswamy (2003) and Xie (2001), total accruals are defined as the difference between income before extraordinary items (data # 18) and net cash flow from operating activities (data # 308). The model is estimated separately for each two-digit SIC industry group within each year.

---

[31] Other studies that have used the S&P T&D ranking include Durnev and Kim (2005) and Cheng, Collins, and Huang (2003).

[32] Jone's cross-sectional model is preferred over its time-series version because it yields a larger sample, enhances the precision of the model, avoids the non-stationary problems in time-series data, and improves the power of tests (Balsam, Chen, and Sankaraguruswamy, 2003).

4.4.2.2 Firm size

One of the oldest proxies used as a measure of information asymmetry is firm size. Vermaelen (1981) and Atiase (1985) interpret firm size as a measure of information asymmetry.[33] These studies suggest that larger firms have more publicly available information about future prospects. Atiase (1985) demonstrate that larger firms will have less information asymmetry before announcements, which is consistent with private pre-disclosure information dissemination is an increasing function of firm size. Given the above, larger firms should be more transparent. This study uses two related measures of firm size, namely, market value of the equity and the number of employees in the firm. Market value of the equity is the product of the common shares outstanding (Compustat annual data # 25) and the year-end closing price (data # 24). Data # 29 gives the number of employees in the firm.

4.4.2.3 Market-to-book ratio

Smith and Watts (1992) find that managers of high-growth firms have superior private information about their firms' cash flow from assets in place as well as investment opportunities. In addition, the nature of growth firms renders them prone to high information asymmetry. This could be a result of either a highly dynamic environment or simply psychological and behavioral factors affecting the investors. Daniel and Titman (2003) argue that investors tend to be more confident about their ability to evaluate information that is relatively vague (such as growth options), and tend to overreact to such intangible information. This tendency is likely to worsen the firm-to-investor information asymmetry. Smith and Watts (1992) use the ratio of market value to book value of assets as a proxy for expected future growth.[34] Intuitively, market value captures the present value of growth opportunities, while book value approximates the value of assets in place. This ratio should be positively related to firm-to-investor information asymmetry.[35] We calculate the market-to-book ratio as:

---

[33] Some of the other studies which use this proxy include Atiase (1985); Bamber (1987); Freeman (1987); Diamond and Verrecchia (1991); Llorente et al. (2002); and Chae (2005).

$$MB = \frac{(Common\ shares\ outs\tan ding)\times(Share\ \Pr ice)+(Total\ assets)-(Common\ equity)}{(Total\ assets)}$$

where Common shares outstanding is obtained from Compustat data # 25, and data # 24 gives the year-end closing share price. Total assets are as given by data # 6, and data # 60 gives the common equity. All data item numbers correspond to the Compustat annual file.

4.4.2.4 R&D to sales ratio

Aboody and Lev (2000) suggest that R&D expenditures are undertaken to generate private information for the firm. Thus, levels of R&D should be positively related to the level of information asymmetry about the firm. The commonly used proxy for the intensity of R&D is the ratio of R&D expenditure (Compustat data # 46) to total sales (data # 12).

---

[34] Some other papers using this measure are Houston and James (1996), and Hegde and McDermott (2004).

[35] This proxy is likely to contain errors in measurement problem arising from issues of values of long-lived assets. Further, Gaver and Gaver (1993) point out that leverage impacts the usefulness of this ratio as a proxy for growth opportunities.

4.4.3 Financial analyst-based proxies

Models of information asymmetry such as the one devised by Miller and Rock (1985) directly link information asymmetry to the firm's earnings. This has led to the development of other proxies of information asymmetry which are more closely related to earnings, for example, the number of analysts providing earnings forecasts for the firm and the dispersion in the provided earnings forecasts.

4.4.3.1 Number of analysts providing earnings forecasts

Fried and Givoly (1982), O'Brien (1988), and Brown (1996) show that financial analysts' earnings forecasts are good measures of the market's expectations. Motivated by these studies, Dempsey (1989), Lobo and Mahmoud (1989), and Coller and Yohn (1997) interpret the number of analysts forecasting the firm's earnings as a measure of information asymmetry. According to these studies, the existence of analyst coverage should reduce information asymmetry. Hong, Lim, and Stein (2000) show that holding all else equal, the more analysts covering a company, the more firm-specific information will be produced and the faster that information will be transmitted. The theoretical support for this measure can be traced to Blackwell and Dubins (1962). They show that opinions tend to converge as the amount of information available about an unknown quantity increases. Thus, *ceteris paribus,* the number of analysts covering a firm should be inversely related to the magnitude of information asymmetry about the firm.

4.4.3.2 Analysts' forecast error

Elton, Gruber, and Gultekin (1984) and Best and Zhang (1993) use financial analysts' earnings forecast errors as a measure of information asymmetry. According to their study, higher errors in analysts' earnings forecasts should be related to larger firm-to investor information asymmetry. Analysts' earnings prediction error (mean forecast - current year earnings) is usually measured as a percentage of the stock price.[36]

4.4.3.3 Coefficient of variation of analysts' forecasts

Another related measure of firm-to-investor information asymmetry is the coefficient of variation (CV) of the forecasts, which is measured as the standard deviation of the current fiscal year earnings forecasts divided by the consensus mean of current fiscal year earnings forecasts. Larger firm-to-investor information asymmetry should be related to a greater variation in the analysts' earnings forecast and, hence, to a higher CV.

4.4.4 Control variables

Boot and Thakor (1993) demonstrate that the incentive for private information acquisition is positively related to financial leverage. This is because increased debt is associated with a greater probability of financial distress. To the extent that this creates valuation uncertainties, greater leverage could be positively associated with information asymmetry. However, as modeled by Ross (1977), greater leverage can signal the quality of a firm, and thus reduce uncertainty. Thus, the association of leverage with information asymmetry is not clear. We use leverage as a control variable to address the concerns raised by Gaver and Gaver (1993). They point out that the usefulness of market-to-book as a measure of firm-to-investor information proxy is affected by leverage. The ratio of long-term debt (Compustat data # 9) to total assets (Compustat data # 6) is used as a measure of financial leverage.

---

[36] Brous (1992), Christie (1987), and Pound (1988) discuss the merits of normalizing by stock price per share, and suggest that normalizing by price results in a better characterization of the importance of the error rather than normalizing by either mean forecasted earnings or actual earnings.

Verrecchia (1983) suggests that managers are likely to provide more informative disclosures when they have good news rather than bad news. This should lead to a relative decrease in information asymmetry in good years, and a relative increase in information asymmetry during loss years. Nevertheless, during a loss year, the manager may seize the opportunity to take a "big bath" and reveal all previously undisclosed bad news at once (Hutton, Miller, and Skinner, 2000). This would lead to a reduction in information asymmetry during the bad years. Thus, the association between profitability and information asymmetry is unclear. In any case, to the extent that it might be affecting firm-to-investor information asymmetry, we control for its effects in this study. A dummy variable is created to control for the profitability effect. The loss dummy takes the value of one if earnings before extra-ordinary items (Compustat data # 18) is less than zero, and is zero otherwise.

Jensen and Meckling (1976) show that as managerial ownership increases, a manager's incentive to exploit outside shareholders decreases. Hence, the manager is less likely to take actions that reduce shareholder wealth. Therefore, high managerial ownership motivates managers to use corporate disclosures in the best interests of shareholders (the interest alignment view of insider ownership). The interest alignment view suggests that insider ownership is associated with greater transparency within a firm. In contrast, Stulz (1988) argues that when manager ownership is relatively small, manager interests might be aligned with outside shareholders but increased ownership can lead to greater agency costs, as managers become more entrenched. Morck, Shleifer, and Vishny (1988) document a non-monotonic relation between manager ownership and firm value, and conclude that the interest alignment and entrenchment views dominate over different ranges of manager ownership. To the extent that insider ownership is related to disclosure, it will also be related to firm-to-investor information asymmetry. We control for the effect of insider ownership using the average per capita insider ownership (percentage of shares held by insiders, divided by the number of insiders holding equity ownership).

While the ratio of R&D expense to sales would be a better measure of information asymmetry between the firm and the market, the use of this variable results in the loss of a large number of data points. We attempt to control for the intangible asset effect by using the intangible dummy (ID). This variable takes the value one if the firm reported R&D expenses, and is zero if the firm did not report R&D expenses. The adverse selection cost of trading is affected by various market microstructure factors. We attempt to control for these effects by limiting the analysis to NYSE firms and introducing order imbalance and trading volume as control variables. The scaled order imbalance is given by:

$$\text{\% Order Imbalance} = \frac{(Total\ Buyer\ Initiated\ trade) - (Total\ Seller\ Initiated\ trade)}{(Total\ Buyer\ Initiated\ trade) + (Total\ Seller\ Initiated\ trade)}$$

4.5 Sample Selection and Sample Characteristics

The sample period runs from 1993 to 2002. Data are retrieved from the NYSE Trade and Quote (TAQ), the Compustat, and the Center for Research in Security Prices (CRSP) databases. Ownership data are retrieved from SEC Compact Disclosures and analyst forecast data was obtained from the IBES database. AIMR scores are from the 1996 edition of the Annual Review of Corporate Reporting Practices by the AIMR. S&P T&D scores are from Patel and Dallas (2002).

Utilities (SIC code 49 to 50) and firms from the financial sector (SIC codes from 60 to 68) are excluded because they are regulated industries. Firms with a fiscal year-end other than December are dropped from the sample. ADRs or other securities, incorporated outside the US, as well as preferred stocks and other non-common stocks, are excluded.[37] All non-NYSE firms are excluded from the sample.[38] All firms involved in any spin-off activity in a given year are excluded from the analysis.[39] To be included in the sample, a firm had to be present in both TAQ and CRSP for all 12 months of the respective fiscal year. To avoid undue influence from extreme observations, firms with stock prices below $5 or above $500 are excluded. Several filters are employed to ensure the validity of the TAQ data.[40] The first trade of each day is dropped from the analysis, since it usually occurs through a call auction. The TAQ database does not eliminate auto-quotes (passive quotes by secondary market dealers), which can cause quoted spreads to be artificially inflated. Since reliable filtering out of auto-quotes in the TAQ is not possible, only the BBO (best bid or offer)-eligible primary market (NYSE) quotes are used.[41] Quotes established before the opening of the market or after the close are discarded. Negative bid-ask spread quotations, negative transaction prices, and negative quoted depths are discarded. Trades with non-standard settlement conditions are excluded.[42] The first trade of each day is discarded to avoid the effects of the opening procedure. Following Lee and Ready (1991), any quote less than five seconds prior to the trade is ignored and the first one at least five seconds prior to the trade is retained.

---

[37] Securities with CRSP share codes other than 10 or 11 were excluded.

[38] The spread decomposition methodologies used in this paper are appropriate for a market-maker (NYSE), as opposed to dealer markets (NASDAQ). In addition, interpretation of the spread components for NASDAQ trade and quotes is problematic due to the presence of inter-dealer trades in the data. These non-information motivated trades cannot be identified in the database. Restricting this study to NYSE-based firms also helps us to control for market microstructure effects on the adverse selection component of the spread.

[39] Excluding spin-offs from the analysis allows us to carry out an out of sample alternate test for the non-monotonic relationship, in Section 7 of this paper.

[40] We drop all trades with a correction indicator other than 0 or 1, and retain only those trades for which the condition is B, J, K, or S. We also drop all trades with a non-positive trade size or price. Finally, we omit all trades recorded before opening time or after the closing time of the market. Negative bid-ask

Table 4.1 presents the distribution of the sample over the period, as well as some descriptive information about the sample firms. The sample size remains fairly stable across the years. The average market capitalization of the sample seems to increase over time, though the trend is not monotonic. The number of employees is used as an alternative measure of firm size. The numbers are very stable across the sample period, unlike the market capitalization, and we do not see any time trend in the number of employees. Average market capitalization is highest in 1999; it shows a monotonic increase from 1993 to 1999, followed by a monotonic decline from 1999 to 2002.

---

spreads and transaction prices are also eliminated. In addition, only quotes that satisfy the following filter conditions are retained: we eliminate all quotes for which (quoted spread is greater than 20% of the quote mid-point, when the quote mid-point is greater than $10) or (quoted spread is greater than $2, when the quote mid-point is less than $10). We also eliminate all quotes for which either the ask- or the bid-quote moves by more than 50%.

[41] All quotes with conditions 5, 7, 8, 9, 11, 13, 14, 15, 16, 17, 19, 20, 27, 28, 29 are excluded.

[42] All trades with conditions A, C, D, N, O, R, or Z are excluded.

Financial leverage shows a marginally increasing trend from 1993 to 1999, followed by a marginal decline from 1999 to 2002.

Table 4.2 (Panel A) presents the pooled descriptive statistics for the firm characteristics variables. The mean (median) level of R&D expense is 3.27% (1.91%) of net sales. As expected, dispersion among firms, with respect to their level of R&D expenditure, is high (4.16%). Leverage varies from 0 to 0.89, with an average (median) level of 0.243 (0.229). Raytech Corporation, with a market capitalization of $7.7 million (year-end, 2000) is the smallest member of the sample, while GE, with a market of $507 billion (year-end, 1999) is the largest firm in the sample. The AIMR score column presents the percentile scores for the disclosure quality of the sample firms from 1993 to 1995. The mean and the median are close to 50, while the minimum is at zero and the maximum is at 100. This suggests that the selected sample has a balanced representation of both opaque and transparent firms. The S&P T&D score presents decile scores for the disclosure quality (annual report basis) of the sample firms in year 2002. The mean T&D score is 4.76 and a median score is 5. The firm with the best disclosure quality has a score of 8 while the worst firm in the sample has a score of 1. Both the mean and median of the discretionary accruals (DAC) are positive: approximately 3% of lagged assets, indicating that managers are more likely to use their discretion to increase earnings.

Table 4.2 (Panel B) presents the pooled descriptive statistics for the analysts' variables and the ownership variables used in this study. The number of analysts providing earnings forecasts for the various NYSE firms ranges from 0 to 47. An average (median) of 12 (10) analysts provide earnings forecasts for each firm in the sample. Our sample firms contain on average (median) 20 (19) insider shareholders. The average (median) insider shareholding is 11.5% (3%).

Table 4.2 (Panel C) presents the pooled descriptive statistics for the market variables. We compute the adverse selection component of the bid-ask spreads for every sample in the firm, for every year in the sample. A year is defined as the period from January 1st to December 31st. This also corresponds to the fiscal years of the sample firm. λ is the computed spread component, and is a fraction of the quoted spread. Approximately 29% of the quoted spread for the typical sample firm is due to the adverse selection problem faced by the market-maker. Dval_GH is the dollar value transformation of λ, which is obtained using the relation:

$$Dval\_GH = \left(\lambda / Average\ trading\ price \right) \times \left(Average\ spread \right)$$

The mean adverse selection cost of trading $100 in the basket of sample firms is about 24 cents while the median is about 16 cents. The average spread is the monthly average of the bid-ask spreads. Both the mean and the median spreads for the typical sample firm are about 15 cents. Order imbalance (number of buyer-initiated trades minus the number of seller-initiated trades) varies from a low of -60% to a high of 64.4% of the total number of trades. The average (median) imbalance is about 12.7% (14%) per year.

The NYSE TAQ data for the ten-year period from January 1993 through December 2002 is used to generate the Glosten and Harris adverse selection component of spread. The period contains 728,709,698 trade observations in total. The number of observations per security ranges from a maximum of 1,693,056 to a minimum of 1,344 observations per year.

By convention, the estimates of the adverse selection cost of trading (λ) are usually scaled by the average trading price. Nevertheless, in the next section, non-scaled λ is used in the regression analysis because market capitalization and average analysts' earnings forecast errors, scaled by price, are present in the models as explanatory variables. Scaling λ by price will potentially obscure the association between the adverse selection component of the spread and other proxies of information asymmetry.

4.6 Empirical Analysis and Results

Figure 1 displays the cross-sectional mean level of inter-investor information asymmetry, expressed in terms of dollar cost per $100 traded.[43] Barring the slight increase in adverse selection cost of trading during 1993 and 1998, we find that the level of inter-investor information asymmetry has been declining consistently, which is not surprising. Some research (see Bollen and Whaley, 1998; Goldstein and Kavajecz, 2000; Jones and Lipson, 2001) has reported declines in spreads and depths following the conversion of trading in eighths to sixteenths on the NYSE in June 1997. Another possible explanation for the declining adverse selection problem faced by the market-maker could be the improved technology that may have led to faster and better information dissemination. Regulation fair disclosure (October 2000) and the Sarbanes-Oxley Act (July 2002) may also be partly responsible for the declining average adverse selection cost of trading in the market.

---

[43] The numbers can be also viewed as cents per dollar traded.

Figure 2 presents time-series patterns in inter-investor information asymmetry and firm-to-investor information asymmetry across industries.[44] This figure compares the relation between the two types of information asymmetries for the set of firms constituting the S&P 500 index in 2002 (October). Inter-investor information asymmetry is expressed in dollar cost per $100 traded. Firm-to-investor information asymmetry is the average S&P T&D score.

In Figure 2, an interesting pattern emerges, where high technology industries such as drugs, genetic engineering, and computers have the lowest levels of inter-investor information asymmetry, while industries such as retailers and utilities display relatively higher levels. As expected, the drugs, genetic engineering and computer industries have relatively high levels of firm-to-investor information asymmetry level. The wholesalers' industry has the lowest level of firm-to-investor information asymmetry. The pattern observed in Figure 2 suggests that measures of firm-to-investor information asymmetry and inter-investor information asymmetry are not synonymous.

We begin by exploring the relation between adverse selection costs and various proxies of firm-to-market information asymmetry, using a simple pair-wise correlation and present some univariate analysis results. The multivariate analysis proceeds with an OLS specification.

4.6.1 Univariate analysis

Table 4.3 presents the Spearman's rank correlation matrix of the various proxies of firm-to-investor information asymmetry and the control variables used in this study.

---

[44] We use an adapted version of the 14-industry classification, as proposed by Ritter and Welch (2002).

The pair-wise correlations between the various indirect proxies for firm-to-investor information asymmetry with the two direct measures (S&P T&D ranking and AIMR disclosure ranking) are small. In any case, the signs are generally consistent with the existing literature.

Higher market-to-book ratio implies more growth options, and thus, higher firm-to-investor information asymmetry. This ratio is negatively correlated with both the S&P T&D ranking and the AIMR ranks. Higher ranks on both the scales correspond to better quality of disclosure and higher transparency, and therefore, to a lower firm-to investor information asymmetry. The absolute level of discretionary accruals (DAC) and the level of expenditure (per dollar sales) in R&D activities (RND) are both negatively correlated with both AIMR and S&P scores. These scores are also negatively correlated with the two analyst-based proxies of firm-to-investor information asymmetry. Analysts' earnings prediction errors (EPE) and dispersion in analysts' forecasts (CV) should be higher for more opaque firms. Larger firms are generally more transparent.

The negative correlation between the market-to-book ratio and the two analyst-based measures seem to be rather counterintuitive; however, to the extent that these proxies are less than perfect measures, some of the between-proxy correlations could be driven by some other firm characteristics. For example, the positive 0.326 correlation between size (market capitalization) and the market-to-book ratio is seemingly driven by the presence of the market value in both measures.

Figures 3A through 3F present the relation between the inter-investor information asymmetry and the various proxies of firm-to-investor information asymmetry. The former is expressed in terms of dollar adverse selection cost per $100 traded. The x-axis represents the decile groups of firm-to-investor information asymmetry. The sample is divided into deciles, based on each of the indirect measures (CV, EPE, DAC, and MB). The groups are created on a yearly basis to remove any calendar effects. The S&P ranking is on a scale of one to ten and, hence, does not require any transformation. The AIMR percentile ranks are collapsed into deciles by grouping together intervals of ten percentiles.

Consistent with the argument laid out in Section 4.2, we find that the relation between inter-investor IA and firm-to-investor IA is concave from below. λ is highest at some point between the 1st and the 10th deciles. The adverse selection cost of trading corresponding to the S&P T&D measure and attains a maximum of about 4.3 cents per $100 traded for firms ranked 4. The lowest point is about 2.8 cents for firms ranked 8. The data does not contain any firms ranked higher than 8, and thus, represents the set of the most transparent firms in the sample. The most opaque firms (rank 1), and firms in the 7th decile, have equal adverse selection cost priced into their spreads (2.99 cents).

Figure 3B presents the adverse selection cost of trading, which corresponds to the various transparency deciles calculated using the AIMR percentile ranks. The lowest adverse selection cost per $100 traded is incurred by traders in the set of the most transparent firms (about 8.1 cents for the 10th decile firms). Nevertheless, the highest cost is incurred by the 4th decile firms' traders (19.4 cents). The adverse selection cost of trading in the set of the most opaque firms (1st decile) is about 10 cents. A similar inference can be drawn from the two analyst-based measures (Figures 3C and 3D), the levels of discretionary accruals (Figure 3E), and the level of market-to-book ratio (Figure

3F). The level of inter-investor information asymmetry for the S&P T&D data is significantly lower than the AIMR data. This could possibly be because the S&P T&D data are for the firms in the S&P 500 index, while the AIMR data also includes non-S&P 500 firms. On average, S&P firms are likely to be more liquid and have lower adverse selection problems. Another possible reason can be inferred from Figure 1. The AIMR data corresponds to years 1993 through 1995, while the S&P T&D data is for year 2002. The level of λ in 2002 is significantly lower than at any point in the years 1993 through 1995.

4.6.2 Box-Cox transformations

We use the Box-Cox transformations described in Section 4.2 to estimate the optimal functional form of the relation between each of the two direct measures of firm-to-investor information asymmetry (decile S&P T&D ranks, and the percentile AIMR ranks) as the independent (x) variable, and the inter-investor information asymmetry (dollar per $100 traded) as the dependent (y) variable. For both measures of firm-to-investor IA, the first stage analysis presents two possible transformations: θ=1 (linear term) and θ=0 (logarithmic transformation). We use both of the suggested transformations to construct the following functional form:

$$\lambda = \beta_0 + \beta_1 \times x + \beta_2 \times \ln(x) + \varepsilon \qquad (4.8)$$

We use the maximum likelihood method to estimate β_0, β_1, and β_2. Figures 4A and 4B present the estimated models. For the S&P T&D firms, the function attains a maximum at x equal to 3.39. At this level of transparency, 3.79 cents worth of adverse selection cost is incurred for every $100 traded. The cost for the most transparent firm (rank 10) is about 1 cent, while the cost for firms in group 1 is about 2.1 cents. *Ceteris paribus,* the adverse selection cost of trading increases as firms become more transparent until it reaches 3.39, at which point, it steadily declines. The AIMR scores attain maxima at 25.92 (15.26 cents for $100 of trade). As firms become more opaque (rank less than 25.92), the cost decreases; similarly, as the firms become more transparent, the adverse selection cost declines. The cost of trading the most transparent basket (rank 100) is about 8.45 cents per $100 of trade. Traders of rank 1 (opaque) firms incur about 5 cents adverse selection cost for every $100 worth of trade.

4.6.3 Univariate regression analysis

Using simple polynomial regression (Equation 4.4), Section 4.6.2 explores the univariate, curvilinear association between the firm-to-market information asymmetry and the inter-investor information asymmetry, as measured by the adverse selection cost component of the spread. The model estimated in this section is of the form:

$$\lambda_{t,j} = \alpha_t + \beta_{1,t} \times IA_{t,j} + \beta_{2,t} \times IA_{t,j}^2 + \varepsilon_{t,j} \qquad (4.9)$$

where, $\lambda_{t,j}$ is the adverse selection cost of trading firm j's shares in year t. The adverse selection cost is expressed in dollars per $100 trade. IA is the measure of the firm-to-investor information asymmetry. According to the hypothesis discussed in Section 4.2, we expect the relation between the firm-to-investor information asymmetry (IA) and the inter-investor information asymmetry (λ) to be non-monotonic. We expect λ to be highest for firms that are neither very opaque (high IA) nor very transparent (low IA). Therefore, we expect β_1 to be positive and β_2 to be negative.

Table 4.4 presents time-series averages of the estimated coefficients. The levels of significance correspond to t-tests for mean equal to zero. The first-order derivative $(\beta_1 + \beta_2 \times IA)$ gives the rate of change in inter-investor IA corresponding to a change in firm-to-investor IA. We find β_1 to be positive and β_2 to be negative for all measures of IA. The results suggest that as firms become more opaque (IA increases) λ increases (positive β_1) because increased opacity provides more opportunities for smart investors to derive greater benefits from firm-specific information. Nevertheless, simultaneously increasing search costs eliminate the marginal informed investors. This increases the proportion of uninformed traders in the market, which causes λ to decline (negative β_2).

4.6.4 Multivariate regression analysis

Table 4.5 presents the results of multivariate regression analysis. The model estimated in this section is of the form:

$$\begin{aligned}\lambda_{t,j} = {} & \alpha_t + \beta_{1,t} \times MB_{t,j} + \beta_{2,t} \times MB^2_{t,j} + \beta_{3,t} \times \ln(Anal)_{t,j} + \beta_{4,t} \times (\ln(Anal))^2_{t,j} + \\ & \beta_{5,t} \times CV_{t,j} + \beta_{6,t} \times CV^2_{t,j} + \beta_{7,t} \times EPE_{t,j} + \beta_{8,t} \times EPE^2_{t,j} + \beta_{9,t} \times DAC_{t,j} + \beta_{10,t} \times DAC^2_{t,j} + \\ & \beta_{11,t} \times AIMR_{t,j} + \beta_{12,t} \times AIMR^2_{t,j} + \beta_{13,t} \times SPTD_{t,j} + \beta_{14,t} \times SPTD^2_{t,j} + \beta_{15,t} \times \ln(Emp)_{t,j} + \\ & \beta_{16,t} \times (\ln(Emp))^2_{t,j} + \gamma_t \times (\text{Various Control variables}) + \varepsilon_{t,j}\end{aligned} \tag{4.10}$$

where, $\lambda_{t,j}$ is the adverse selection cost of trading firm j's shares in year t. Unlike model (9), where the adverse selection cost of trading is expressed in dollar terms, this model uses adverse selection cost expressed as percentage of the spread.[45] $MB_{t,j}$ is the market to book ratio of firm j in year t; $MB_{t,j}^2$ is the square of the market to book ratio; $\ln(\text{Anal})_{t,j}$ is the natural logarithm of the number of institutional analysts providing forecasts for firm j in year t; CV is the coefficient of variation in the analysts' forecasts; and EPE is the error in the analysts' forecasts. DAC is the level of performance-adjusted discretionary accruals; AIMR and SPTD are the disclosure quality scores obtained from the AIMR database, and the S&P T&D database, respectively. $\text{Ln}(\text{EMP})_{t,j}$ is the natural logarithm of the number of employees in firm j in year t. The control variables include PC_INSDR (percentage of shares held by insiders, divided by the number of insiders holding shares), leverage (lev), trading volume (ln(Vol)), scaled order imbalance (OI), loss dummy (LD), and intangible dummy (ID).

We estimate various reduced forms of model 10 as a cross sectional model in every year. Table 4.5 presents the time-series averages of the estimated coefficients. The levels of significance represent the result of t-tests for mean equal to zero. The adjusted R^2 presents the time-series average adjusted R^2. The multivariate analysis results are identical to the results for the univariate analysis. This is not surprising, given the low level of pair-wise correlations between the variables. A discussion of the results follows.

4.6.4.1 Market to book ratio (MB)

Market-to-book ratio measures the growth opportunity of a firm relative to the value of the assets in place. This ratio should be positively related to information asymmetry between the firm/management and the outside investor. The market-to-book ratio of the sample firms ranges from 0.43 to 6.96 (Table 4.2). The estimated model suggests that for values from 0.43 to 4, the market-to-book ratio is positively related to the adverse selection cost of trading, while, in the range from 4 to 6.96, this ratio is negatively associated with λ (adverse selection cost of trading). This result concurs with the hypothesis in Section 4.2. Thus, the book-to-market ratio (which is a measure of opacity of a firm) is associated with λ by a non-monotonic relation, and is concave from below (Table 4.5).

---

[45] Using dollar λ instead of % λ does not change the results. We present the results corresponding to % λ as the independent variable. Several of the independent variables in the model are correlated with price. Dividing % λ by price (to get dollar λ) could obscure the relationship under investigation.

4.6.4.2 Number of analysts providing earnings forecasts (LnAnal)

LnAnal is the natural logarithm of the number of analysts that cover the firm (as reported in IBES). Brennan and Subrahmanyam (1995) find that the number of analysts following a firm is directly related to the amount of produced information. Thus, the larger the number of analysts following a firm, the lower will be the information asymmetry about the firm. Brennan, Jegadeesh, and Swaminathan (1993) find that the stock price of firms that have a greater number of analysts following tends to react more rapidly to new information than does the stock price of firms with fewer analysts.

The results in Table 4.4 suggest that, as the number of analysts following a firm increases from zero to three, the adverse selection cost in the market rises marginally. However, beyond this point, as the number of analysts increases, λ shows a very fast decline towards a new low point. Table 4.5 predicts that if one controls for other firm characteristics, the zero-to-three interval widens to as much as zero to fifteen. We note that when controlling for liquidity, leverage, size, insider ownership, earnings losses, and presence of intangibles in the firm, the adverse selection cost in the market rises as the number of analysts following the firm increases from zero to fifteen. A possible explanation for this could be that we are controlling for most common information sources used by the analysts. Holding these sources constant reduces the per capita volume of useful information generated by the analysts. The effect of this is similar to reducing the number of analysts in the uncontrolled setup.

The observed result supports the predicted curvilinear relation, whereby, small number of analysts generates private information accessible to a small number of investors. This leads to higher adverse selection risk for the uninformed agents in the market. As the number of analysts increases, the information they generate is available to a relatively larger number of investors. This leads to the faster incorporation of the private information into the stock prices, leading to a decline in λ.

4.6.4.3 Coefficient of variation (CV) and errors (EPE) in analysts' forecasts

The coefficient of variation (CV) is a measure of the quality of the analysts' forecasts. Controlling for the number of analysts, the coefficient of variation of their earnings forecasts is a measure of both the information asymmetry about the firm and the rate at which the produced information is incorporated into the price process. First, more opaque firms will be more difficult to analyze, and hence, have a greater likelihood of more dispersed earnings forecasts. Second, a lower dispersion in analysts' forecasts will imply a less noisy signal, and hence, a faster incorporation of the information into the price process. To the extent that CV is a measure of the noisiness of the firm's signals, its association with adverse selection risk in the market will be similar to the association between the market-to-book ratio and the adverse selection risk in the market. Similarly, the magnitude of earnings prediction errors (EPE) captures the level of difficulty faced by analysts in generating earnings forecasts for a given firm. We expect this level of difficulty to be higher for opaque firms. Therefore, EPE may be interpreted as a proxy for the firm-to-investor information asymmetry. The results in Table 4.5 support the hypothesized non-monotonic relation between firm-to-investor IA (with CV and EPE as proxies) and inter-investor IA.

4.6.4.4 Discretionary accruals (DAC)

DAC captures the extent to which management records non-cash income or expense items in an aggressive (income increasing) or conservative (income decreasing) manner. The recording of discretionary accruals might or might not adhere to generally accepted accounting principles (GAAP). In other words, they can represent management's use of the latitude in GAAP or failure to adhere to GAAP. Therefore, high levels of positive or negative discretionary accruals are often interpreted in the accounting literature as indicators of earnings management. Prior research indicates that analysts have more difficulty forecasting the earnings of firms with high levels of accruals (Bradshaw and Sloan, 2001). Therefore, absolute levels of DAC may be interpreted as a proxy for a firm's transparency.

We find some evidence of a curvilinear relation between DAC and inter-investor information asymmetry. This relation becomes stronger in models 4 and 5 (Table 4.5). These models control for the quality of disclosure using the AIMR disclosure score (Model 4) and the S&P transparency and disclosure index (Model 5). Other disclosures such as notes in the annual reports can go a long way in alleviating accruals- related transparency. These results suggest that levels of accruals are likely to be a more accurate measure of firm-to-investor information asymmetry, controlling for the quality of disclosure.

4.7 An Alternative Test for the Non-Monotonic Relation between Inter-investor and Firm-to-Investor Information Asymmetry

The results discussed so far are based on the analysis of the selected sample. Both the univariate and the multivariate analyses predict curvilinear relations between firm-to-investor information asymmetry and inter-investor information asymmetry. In this section, we test the robustness of our results by analyzing the event month change in inter-investor information asymmetry for a set of firms that have spin-off units. Our spin-off sample consists of 77 firms for which the financial year is from January to December.[46] Following Huson and MacKinnon (2003), we categorize spin-offs into focus-enhancing and non-focus-enhancing. We classify spin-offs as focus-enhancing if the parent and subsidiary have different two-digit SIC codes. All other spin-offs are classified as non-focus-enhancing. We find that 56 out of the 77 spin-offs belong to the focus-enhancing group. We obtain disclosure quality information for 19 of these 56 firms from the AIMR files and obtain analysts' forecasts for all 56 firms from the IBES tapes.

Using the univariate analysis (Table 4.4), we identify the point of maximum for the relation between the AIMR scores and the adverse selection cost of trading in the market. The relation is an increasing function until approximately rank 40, at which point, it starts to decline. We classify the 19 spin-off parent firms as transparent if their AIMR ranks are greater than 40; otherwise we classify the firms as opaque firms.[47] We identify 12 opaque and 7 transparent firms in the sample. Similarly, we identify the breakpoint for the coefficient of variation (CV) measure of firm-to-investor information asymmetry. We classify all firms with a CV less than or equal to 0.18 as transparent, and all other firms as opaque. This criterion allows us to identify 42 of the spin-off parent firms as opaque and 14 as transparent. We identify event month zero as the month of the spin-off. Table 4.6 presents a time-series profile of the dollar adverse selection costs, adverse selection costs as a proportion of the spreads, and the mean quoted spreads for the two sets of transparent and opaque firms. The analysis starts two months prior to the spin-off and ends three months post-spin-off. Based on the argument laid out in this paper, we expect to see an increase in the inter-investor information asymmetry for the set of opaque firms and a corresponding decline for the set of transparent firms.

---

[46] We find 158 spin-offs from SDC platinum between January 1993 and December 2002. Of these, we retain only those firms that have their financial year starting in January and ending in December. This criterion allows us to keep this analysis in sync with the rest of the paper.

We find that opaque firms have higher spreads than do the transparent firms. Surprisingly, the spin-off does not seem to have any clear effect on the overall spread. Nevertheless, we find that the dollar adverse selection cost for transparent firms declines post-spin-off, while the corresponding cost for the set of opaque firms increases. For the AIMR sample, the costs decrease by about 3% in the event month and by about 7% over three months. The costs for the set of opaque firms increases by approximately 12% in the event month and by 18% over three months. From Panel B, using the CV measure, we find a more symmetric result. For the transparent set, the dollar adverse selection cost declines by about 32% during the event month; however, the costs increase slightly over the following 3 months to 25% below the pre-spin-off level. For the set of opaque firms, the costs increase by about 39.25% during the spin-off month. The 3-month change for this group is about 28% above the pre-spin-off level. Since a relatively larger proportion of the sample belongs to the group of opaque firms, this group dominates the net effect. On average, spinoff leads to an increase in the adverse selection cost of trading for the parent firm.

---

[47] All of the 158 spin-off firms are excluded from the analyses in Sections 5 and 6. This allows us to use the break points from Table 4.4 without concern for data mining.

4.8 Conclusions

This paper examines the relation between opacity of a firm (firm-to-investor information asymmetry) and the adverse selection cost of trading its shares (inter-investor information asymmetry). We provide evidence showing that firm-to-investor information asymmetry and inter-investor information asymmetry display a non-monotonic relation. This result holds even after controlling for liquidity, other market microstructure effects, level of debt, level of intangible assets, and other firm characteristics.

The relation between the two types of information asymmetry is determined by the interplay of information search cost and gains derived by trading with superior information. While the former is likely to dissuade informed trading, which would lead to a decline in inter-investor information asymmetry, the latter is likely to encourage informed trading, which, in turn, would lead to an increase in inter-investor information asymmetry. Our results suggest that as very transparent firms become more opaque, the benefits of superior information exceeds the information search costs, leading to an increase in informed trading and, therefore, an increase in inter-investor information asymmetry. Nevertheless, beyond a point, the search cost seems to dominate the derived benefits, leading to a decline in informed trading, and consequently, a decline in inter-investor information asymmetry.

In conclusion, the finding of this study suggests that firm-to-market information asymmetry differs fundamentally from inter-investor information asymmetry. While low inter-investor information asymmetry is desirable for better functioning of financial markets, attempts to achieve it by marginally reducing firm-to-investor information asymmetry might produce the opposite result.

References

Aboody, D., and Lev, B., 2000. Information asymmetry, R&D, and insider gains. *Journal of Finance*, 55(6): 2747-2766.

Admati, A.R., and Pfleiderer, P., 1988. A theory of intraday patterns: Volume and price variability. *Review of Financial Studies*, 1(1): 3-40.

Amemiya, T., and Powell, J. L., 1981. A comparison of the box–cox maximum likelihood estimator and the non-linear two-stage least squares estimator, *Journal of Econometrics* 17: 351–381

Amihud, Y., and Mendelson, H., 1988. Liquidity and asset prices - financial management implications. *Financial Management*, 17(1): 5-15.

Amir, E., and Lev, B., 1996. Value-relevance of non-financial information: The wireless communications industry. *Journal of Accounting and Economics*, 22(1-3): 3.

Atiase, R.K., 1985. Pre-disclosure information, firm capitalization, and security price behavior around earnings announcements. *Journal of Accounting Research*, 23(1): 21-36.

Balsam, S., and H. Chen, S. Sankaraguruswamy, 2003. Earnings management prior to stock option grants. Working paper, Temple University, Georgetown University and Securities and Exchange Commission.

Bamber, L.S., 1987. Unexpected earnings, firm size, and trading volume around quarterly earnings announcements. *Accounting Review*, 62(3): 510-532.

Best, R.J., Best, R.W., and Young A.J., 1998. An examination of proxies of information asymmetry. *Advances in Financial Planning and Forecasting*, 8: 17-34.

Best, R.W, and Zhang, H, 1993. Alternative information sources and the information content of bank loans. *Journal of Finance*, 48(4): 1507-1522.

Bickel, P. J., and Doksitm, K.A., 1981. An analysis of transformations revisited. *Journal of The American Statistical Association,* 76: 296-311.

Blackwell, D., and Dubins, L., 1962. Merging of opinions with increasing information. *The Annals of Mathematical Statistics*, 33(3): 882-886.

Bollen, N.P.B., and Whaley, R.E., 1998. Are ‘teenies’ better? *Journal of Portfolio Management*, 24 (Fall): 10-24.

Boot, A.W.A., and Thakor, A.V, 1993. Security design. *Journal of Finance*, 48(4): 1349-1378.

Botosan, C.A., and Plumlee, M.A., 2002. A re-examination of disclosure level and expected cost of equity capital, *Journal of Accounting Research,* 40(1), 2–40.

Bradshaw, M., Richardson, S., and Sloan, R., 2001. Do analysts and auditors use information in accruals? Journal of Accounting Research 39, 45–74.

Brennan, M.J., Jegadeesh, N., and Swaminathan, B., 1993. Investment analysis and the adjustment of stock prices to common information. *Review of Financial Studies*, 6(4): 799-824.

Brennan, M.J., and Subrahmanyam, A., 1995. Investment analysis and price formation in securities markets. *Journal of Financial Economics*, 38(3): 361-381.

Brous, P.A., 1992. Common stock offerings and earnings expectations: A test of the release of unfavorable information. *Journal of Finance*, 47(4): 1517-1536.

Brown, L.D., 1996. Analyst forecasting errors and their implications for security analysis: An alternative perspective, *Financial Analysts Journal*, 52(1): 40.

Bushee, B.J. and Noe, C.F., 2000. Corporate disclosure practices, institutional investors, and stock return volatility. *Journal of Accounting Research*, 38: 171–202.

Chae, J., 2005. Trading volume, information asymmetry, and timing information. *The Journal of Finance*, 60(1): 413-442.

Cheng, C.S., A. D. Collins, and H. Huang, 2003. The Effect of the S&P T&D Rankings on Market Beta, Abnormal Returns and Earnings Response Coefficients in the Period Surrounding the Report Release Date, Working Paper.

Christie, Andrew A., 1987. On cross-sectional analysis in accounting research. *Journal of Accounting and Economics*, 9(3): 231.

Coller, M., Yohn, T.L., 1997. Management forecasts and information asymmetry: An examination of bid-ask spreads. *Journal of Accounting Research*, 35(2): 181.

Copeland, T. E., and Galai, D., 1983. Information effects on the bid-ask spread. *Journal of Finance*, 38(5): 1457-1469.

Daniel, K. and Titman S., 2003. *Market reactions to tangible and intangible information.* NBER Working Paper 9743

Dempsey, Stephen J., 1989. Predisclosure information search incentives, analyst following, and earnings announcement response. *The Accounting Review*, 64(4): 748.

Demsetz, H., 1968. The cost of transacting. *The Quarterly Journal of Economics*, 82(1): 33-53.

Diamond, D.W., 1985. Optimal release of information by firms. *Journal of Finance*, 40(4): 1071-1094.

Diamond, D.W., and Verrecchia, R.E., 1991. Disclosure, liquidity, and the cost of capital. *Journal of Finance*, 46(4): 1325-1359.

Durnev, A., and E. Kim, 2005. To Steal or Not to Steal: Firm Attributes, Legal Environment and Valuation, Journal of Finance, 60(3): 1461-1493.

Easley, D., and O'Hara, M., 1992. Time and the process of security price adjustment. *Journal of Finance*, 47(2): 577-605.

Easley, D., and O'Hara, M., 2004. Information and the cost of capital. *The Journal of Finance*, 59(4): 1553-1583.

Elton, E.J., Gruber, M.J., and Gultekin, M.N., 1984. Professional expectations: Accuracy and diagnosis of errors. *Journal of Financial and Quantitative Analysis*, 19(4): 351-363.

Freeman, R. N., 1987. The association between accounting earnings and security returns for large and small firms. *Journal of Accounting and Economics*, 9(2): 195-228.

Fried, D., and Givoly, D., 1982. Financial analysts' forecasts of earnings: A better surrogate for market expectations. *Journal of Accounting and Economics*, 4(2): 85-107.

Gaver, J. J., and Gaver, K. M., 1993. Additional evidence on the association between the investment opportunity set and corporate financing, dividend, and compensation policies. *Journal of Accounting and Economics*, 16(1-3): 125-160.

Gelb. D.S. and Zarowin, P., 2002. Corporate disclosure policy and the informativeness of stock prices, *Review of Accounting Studies*, 7: 33-52.

Glosten, L. R., and Harris, L. E., 1988. Estimating the components of the bid ask spread. *Journal of Financial Economics*, 21(1): 123-142.

Glosten, L.R., and Milgrom, P.R., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1): 71-100.

Goldstein M., and Kavajecz, K., 2000. Eighths, sixteenths, and market depth: changes in tick size and liquidity provision on the NYSE, *Journal of Financial Economics* 56: 125–149.

Hakansson, Nils H., 1977. Interim disclosure and public forecasts - an economic analysis and a framework for choice. *The Accounting Review*, 52(2): 396.

Healy, P.M., Hutton, A., and Palepu, K.G., 1999. Stock performance and intermediation changes surrounding sustained increases in disclosure. *Contemporary Accounting Research,* 16 Fall, 485–520.

Hegde, S.P., and McDermott, J.B, 2004. Firm characteristics as cross sectional determinants of adverse selection. *Journal of Business Finance and Accounting*, 31(7-8): 1097-1124.

Ho, T., and Stoll, H.R., 1981. Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics*, 9(1): 47-73.

Hong, H., Lim, T., and Stein, J.C., 2000. Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *Journal of Finance*, 55(1): 265-295.

Houston, J., and James, C., 1996. Bank information monopolies and the mix of private and public debt claims. *Journal of Finance*, 51(5): 1863-1889.

Huson, Mark R., and MacKinnon, G., 2003. Corporate spinoffs and information asymmetry between investors. *Journal of Corporate Finance*, 9(4): 481-503.

Hutton, Amy P., Gregory S. Miller, and Douglas J. Skinner, 2000. "Effective Voluntary Disclosure", Working Paper.

Jensen, Michael C., and William H. Meckling, 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, Vol. 3: 303-360.

Jones C., and Lipson, M., 2001. Sixteenth: direct evidence on institutional execution costs, *Journal of Financial Economics*, 59: 253–278.

Kandel, E., and Pearson N.D., 1995. Differential interpretation of public signals and trade in speculative markets. *Journal of Political Economy*, 103(4): 831-872

Kim, O., and Verrecchia, R.E., 1994. Market liquidity and volume around earnings announcements. *Journal of Accounting and Economics*, 17(1-2): 41-67.

Kim, O., and Verrecchia, R.E., 1997. Pre-announcement and event-period private information. *Journal of Accounting and Economics*, 24(3): 395-419.

Kim, O., and Verrecchia, R.E., 2001. The relation among disclosure, returns, and trading volume information. *The Accounting Review*, 76(4): 633.

Krinsky, I., and Lee, J., 1996. Earnings announcements and the components of the bid-ask spread. *Journal of Finance* 51, 1523-1535

Kyle, A.S., 1985. Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, 53(6): 1315-1336.

Lang, M.H., and Lundholm, R.J., 1996. Corporate disclosure policy and analyst behavior. *The Accounting Review*, 71(4): 467.

Lee, C.M.C., and Ready, M.J., 1991. Inferring trade direction from intraday data. *Journal of Finance*, 46(2): 733-746.

Lee, C.M.C., and Radhakrishna, B., 2000. Inferring investor behavior: Evidence from TORQ data. *Journal of Financial Markets*, 3(2): 83-111.

Lin, J., Sanger, G., and Booth, G., 1995. Trade size and components of the bid-ask spreads. *Review of Financial Studies* 8, 1153-1183.

Llorente, G., M., Saar, R.G., and Wang, J., 2002. Dynamic volume-return relation of individual stocks. *Review of Financial Studies*, 15(4): 1005-1047.

Lobo, Gerald J., and Mahmoud, Amal W.A., 1989. Relation between differential amounts of prior information and security return variability. *Journal of Accounting Research*, 27(1): 116.

Lundholm, R.J., 1991. Public signals and the equilibrium allocation of private information. *Journal of Accounting Research* 29, 322-349.

Lundholm, R., and L. A. Myers, 2002. Bringing the Future Forward: The Effect of Disclosure on the Returns-Earnings Relation. *Journal of Accounting Research*, 40(3): 809-839.

Miller, M.H., and Rock, K., 1985. Dividend policy under asymmetric information. *Journal of Finance*, 40(4): 1031-1051.

Morck, R., Shleifer, A., and Vishny, R.W., 1988. Management ownership and market valuation: An empirical analysis. *Journal of Financial Economics*, 20: 293-315.

Myers, Stewart C., and Majluf, Nicholas S. 1984. Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics*, 13(2): 187-221.

Neal, R., and Wheatley S., 1998. Adverse selection and bid-ask spreads: Evidence from closed-end funds, Journal of financial markets. 1(1): 121-149

O'Brien, P.C., 1988. Analysts' forecasts as earnings expectations. *Journal of Accounting and Economics*, 10(1): 53-83.

Odders-White, Elizabeth R., 2000. On the occurrence and consequences of inaccurate trade classification. *Journal of Financial Markets*, 3(3): 259-286.

Patel, S.A., and G. Dallas, 2002. Transparency and Disclosure: An Overview of Methodology and Study Results – United States, Standard & Poor's, http://governance.standardandpoors.com. pp.1-29.

Pound, John, 1988. The information effects of takeover bids and resistance. *Journal of Financial Economics*, 22(2): 207.

Ritter, J.R., and Welch I., 2002. A review of IPO activity, pricing, and allocations. *Journal of Finance*, 57, 1795-1828.

Ross, S.A., 1977. The determination of financial structure: The incentive-signaling approach. *The Bell Journal of Economics*, 8(1): 23-40.

Smith, C.W., and Watts, R.L. 1992. The investment opportunity set and corporate financing, dividend, and compensation policies. *Journal of Financial Economics*, 32(3): 263-292.

Stulz, R.M., 1988. Managerial control of voting rights: Financing policies and the market for corporate control. *Journal of Financial Economics*, 20: 25-54.

Venkatesh, P.C., and Chiang, R., 1986. Information asymmetry and the dealer's bid-ask spread: A case study of earnings and dividend announcements. *Journal of Finance*, 41(5): 1089-1102.

Vermaelen, T., 1981. Common stock repurchases and market signaling: An empirical study. *Journal of Financial Economics*, 9(2): 139-183.

Verrecchia, R.E., 1983. Discretionary disclosure. *Journal of Accounting & Economics*, 5(3): 179.

Appendix

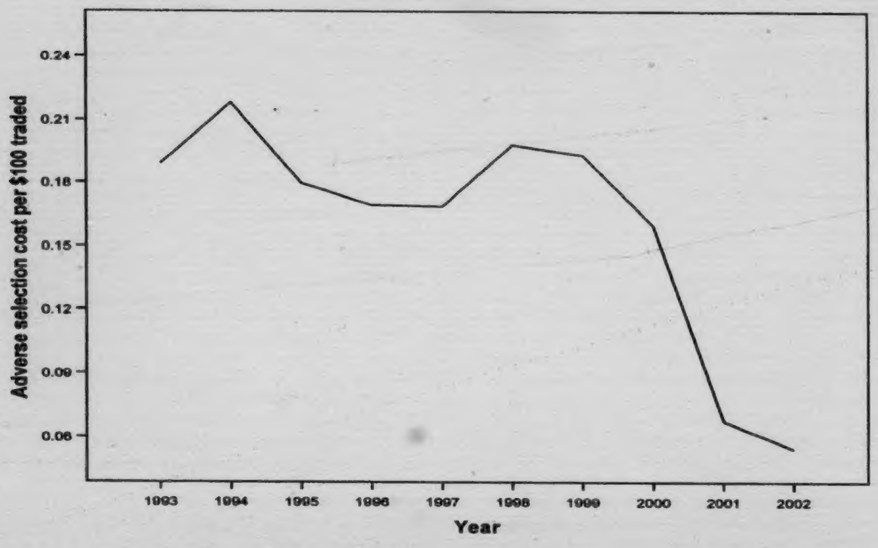

Figure 4.1: Annual average level of the adverse selection cost per $100 traded.

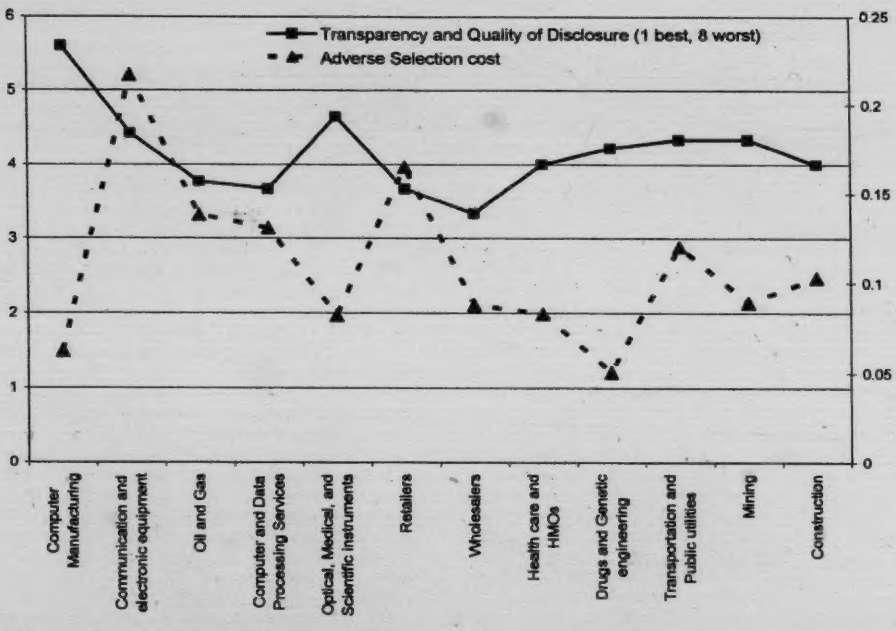

Figure 4.2: Average adverse selection cost of trading, per $100 traded vs. quality of the firm's disclosure (2002 S&P disclosure ranks).

The S&P T&D scores rank firms on a scale of 1 to 8, with worst firms ranked 1 and the best ranked 8. For purpose of comparison, the scores are reversed in the figure below (1 is best and 8 is worst).

Figure 4.3: Observed association between λ and the various proxies of firm-to-market information asymmetry.

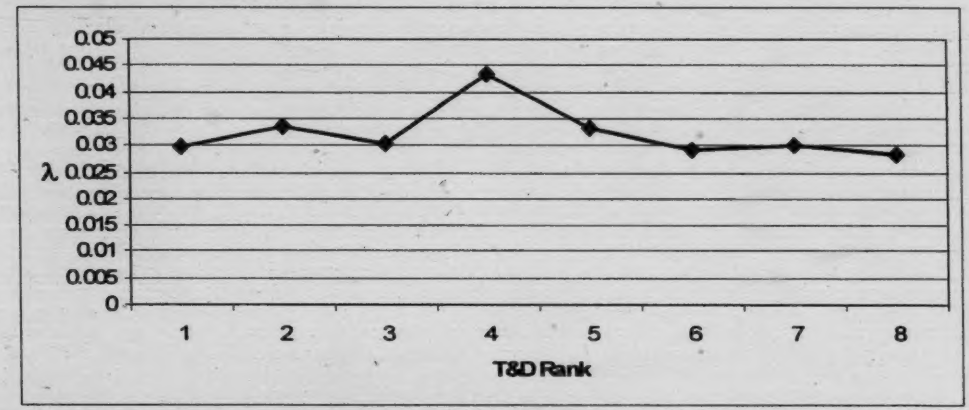

Figure 4.3(A): S&P T&D Scores.

X axis denotes the transparency and disclosure (T&D) ranking as stated in S&P 2002 study. Higher rank implies better quality disclosure and more transparent firm. Y axis denotes the average adverse selection cost per $100 traded.

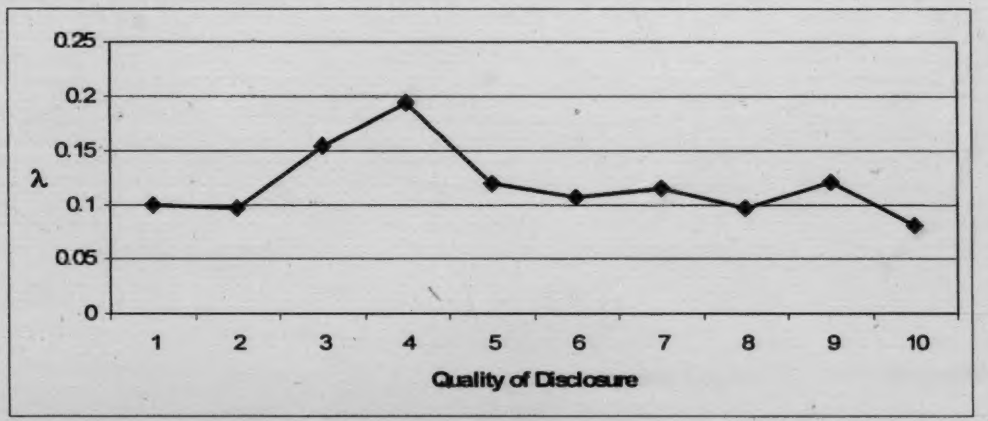

Figure 4.3(B): AIMR disclosure quality rank.

X axis denotes the Disclosure quality rank as stated in AIMR 1993 to 1995 report. Higher rank implies better quality disclosure thereby more transparent firm. AIMR study ranked firms on scale of 1 to 100. In the figure below, 1 denotes all firms with ranks 1 to 10, 2 denotes 11 to 20 and so on. Y axis denotes the average adverse selection cost per $100 traded.

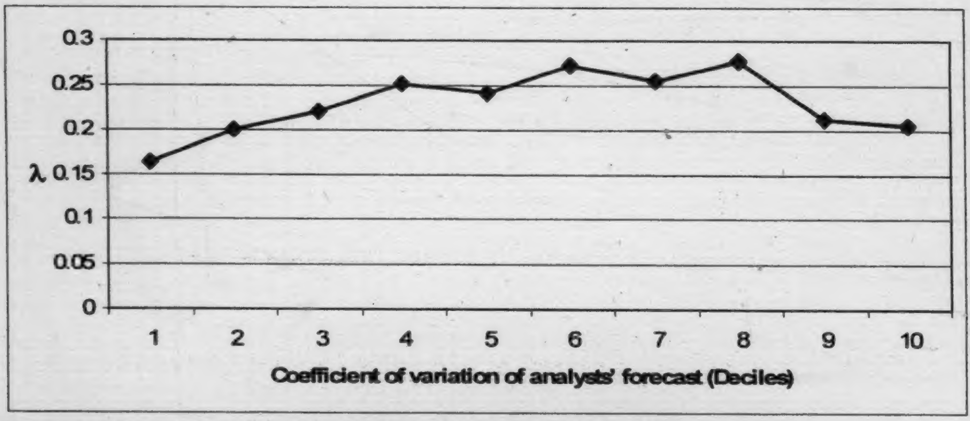

Figure 4.3(C): **Coefficient of variation of analyst forecasts.**

The coefficient of variation of analysts' forecast (CV) is divided into 10 equal class intervals. X axis denotes the class number. Y axis denotes the average adverse selection cost per $100 traded. CV is measured as the standard deviation of current fiscal year earnings forecasts divided by consensus mean of current fiscal year earnings forecasts. We calculate the average λ for each class in each year. The numbers in the figure below are the time-series average calculated across 1993 to 2002.

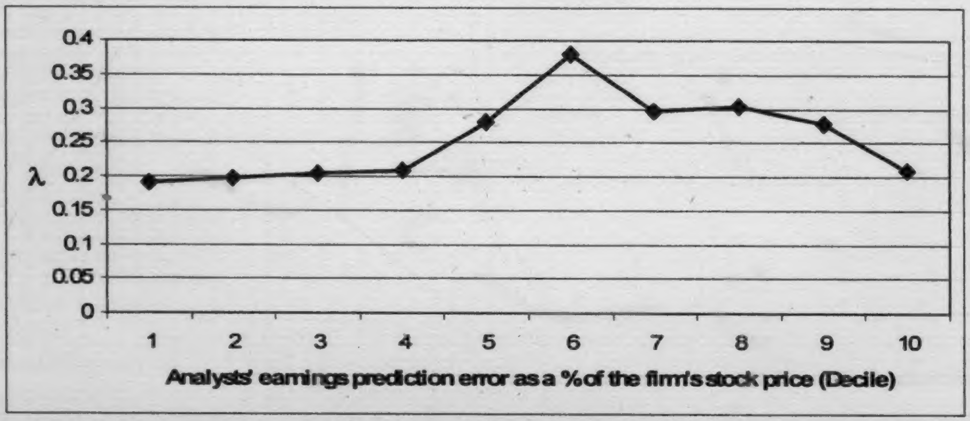

Figure 4.3(D): **Analyst earnings prediction error.**

X axis denotes the decile ranking of Analysts' earnings prediction error, measured as a percentage of the firm's stock price (EPE). Y axis denotes the average adverse selection cost per $100 traded. We calculate the average λ for each EPE rank, in each year. The numbers in the figure below are the time-series average calculated across 1993 to 2002.

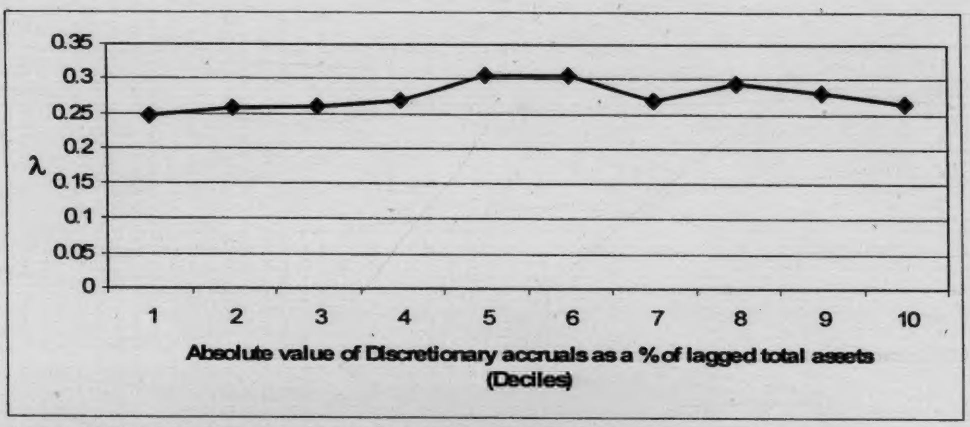

Figure 4.3(E):Performance matched discretionary accruals.

X axis denotes the decile rank of the absolute value of the performance adjusted discretionary accruals (DAC). Higher number implies greater earnings management ability. Y axis denotes the average adverse selection cost per $100 traded. We calculate the average λ for each DAC rank, in each year. The numbers in the figure below are the time-series average calculated across 1993 to 2002.

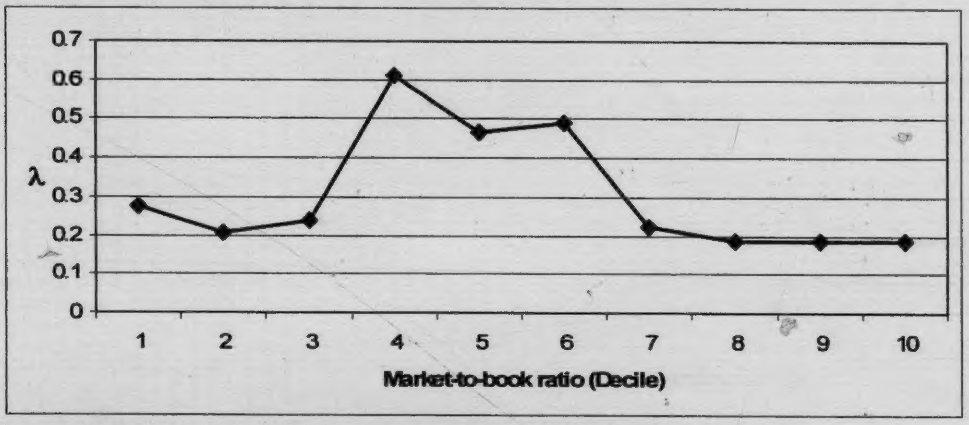

Figure 4.3(F):Market to book ratio.

X axis denotes the decile rank of the market-to-book ratio (MB). Y axis denotes the average adverse selection cost per $100 traded. We calculate the average λ for each MB rank, in each year. The numbers in the figure below are the time-series average calculated across 1993 to 2002.

Figure 4.4: The functional form between two of the firm-to-investor information asymmetry are estimated using the Box-Cox transformation technique.

The y axis represents the $ adverse selection cost per $100 traded. AIMR scores rank firms (1993 to 1995) on a scale of 1 to 100 with 100 being most transparent (best quality disclosure) and 1 being worst. S&P T&D rank S&P 500 firms in year 2002 on a scale of 0 to 10 with 10 being the best and zero the worst.

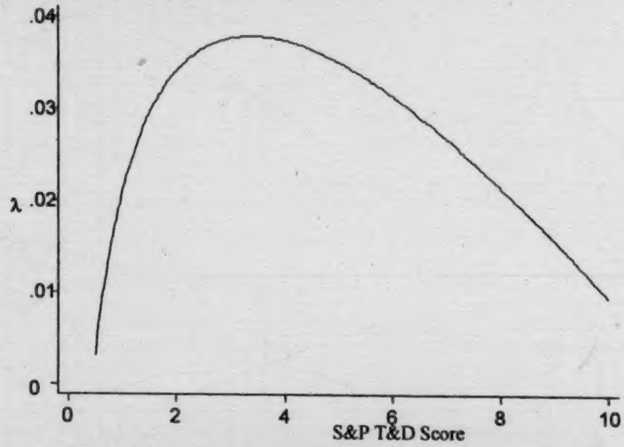

Figure 4.4(A): Box-Cox fitted model for S&P T&D (2002) data.

The fitted model corresponds to the OLS estimated equation:

$$\lambda = 0.030709749 - 0.009654116 \times (\text{S\&P Score}) + 0.032727368 \times \ln(\text{S\&P Score})$$

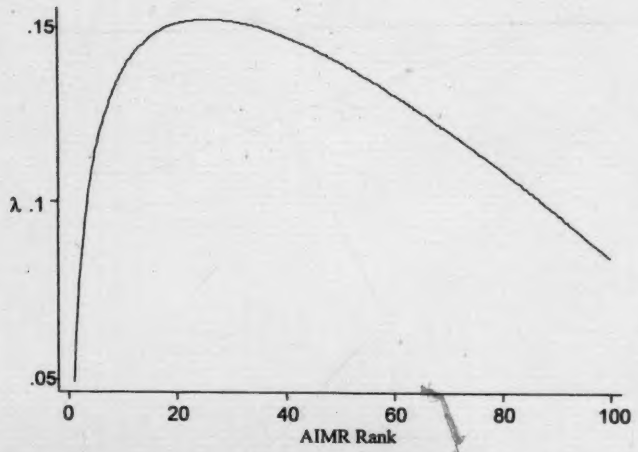

Figure 4.4(B):Box-Cox fitted model for AIMR data.

The fitted model corresponds to the OLS estimated equation:

$$\lambda=0.050819471-0.001741103\times\left(\text{AIMR Score}\right)+0.045130847\times\ln\left(\text{AIMR Score}\right)$$

Table 4.1: Distribution of firms across the sample period. All presented numbers are arithmetic averages.

Year	Number of firms	R&D to Sales ratio	Market-to-Book ratio	Financial Leverage	Size ($000)	Number of Employees
1993	870	0.0343	1.7356	0.2286	3,612,553	22343
1994	918	0.0327	1.6334	0.2278	3,390,099	20870
1995	947	0.0307	1.7784	0.2305	4,449,652	21251
1996	945	0.0303	1.8628	0.238	4,828,017	20315
1997	928	0.0312	1.9814	0.2451	6,155,276	20198
1998	896	0.0345	1.8572	0.2537	7,433,722	20547
1999	839	0.0318	1.775	0.2599	8,537,543	21676
2000	803	0.0323	1.7114	0.247	9,168,130	22686
2001	864	0.0345	1.7486	0.2513	8,358,401	22821
2002	844	0.0361	1.5677	0.2462	6,837,287	22402

Market-to-book, R&D, PPE, and Leverage are obtained from the Compustat annual data tapes.

Size = (Number of shares outstanding) X (Year end closing share price) of Dollars

* All variables are defined in Section 4.4 of the paper.

Table 4.2: **Descriptive Statistics for NYSE-listed sample Firms (1993 to 2002).**

Panel A: Descriptive statistics (Firm Characteristics)

	Market-to-Book ratio	R&D to Sales ratio	Financial Leverage	Size ($000)	Number of Employees	S&P T&D Scores	AIMR Scores	DAC
Mean	1.7729	0.0327	0.2429	6,256,782	21,435	4.76	53.79	0.0362
Median	1.5013	0.0191	0.2291	1,017,308	7,000	5	55.28	0.0294
Std Deviation	0.9261	0.0416	0.1736	21,701,334	45,900	1.05	30.66	0.2079
Maximum	6.9609	0.2895	0.8961	507,216,647	746,000	8	100	0.2132
Minimum	0.4257	0.0000	0.0000	7,706	1,000	1	0	-0.11

Panel B: Descriptive statistics (Analysts and ownership variables)

	Number of Analysts providing forecast	Dispersion in earnings forecast $\left(\sigma_{forecast} / \mu_{forecast}\right)$	Number of Insiders holding equity ownership	Percentage owned by INSIDERS
Mean	11.5	3.5177	20	11.4594
Median	10.0	0.0216	19	2.9000
Std Deviation	8.1	19.5625	13	18.3656
Maximum	47.0	117.1639	119	99.9900
Minimum	1.0	-0.81101	0	0.0000

Panel C: Descriptive statistics (Market variables)

	GH (λ)	Dval_GH (Cents)	Average spread ($)	% Order Imbalance	Share Turnover
Mean	0.2885	0.2418	0.1540	0.1268	0.7860
Median	0.2976	0.1571	0.1543	0.1402	0.6263
Std Deviation	0.1018	0.2661	0.0943	0.1492	0.5960
Maximum	0.9763	3.2120	2.1149	0.6444	3.9446
Minimum	0.0003	0.0003	0.0128	-0.5954	0.0036

Table 4.3: Pearson Correlations between various proxies of information asymmetry.

MB is the market-to book ratio; RND is the reported Research and Development expense, as a percentage of total sales. LEV is the Debt to Asset ratio. Ln(size) is the natural logarithm of the year-end market cap. Ln(Anal) is the natural log of the number of analysts providing earnings forecasts for the firm; EPE is the earning forecast error (Mean forecasted earning – declared earnings), expressed as a proportion of the year-end closing price. CV is the coefficient of variation of earnings forecast. It is defined as standard deviation of current fiscal year earnings forecasts divided by consensus mean of current fiscal year earnings forecasts. ln(INSDR) is the number of insiders holding equity ownership in the firm, and PINSDR is the proportion of shares held by the insider shareholders.

	MB	S&P T&D Score	AIMR Score	DAC	RND	LEV	Ln(Size)	Ln(Anal)	EPE	CV	ln(Insdr)	PINSDR
MB	1											
S&P T&D Score (2002 only)	-0.126	1										
AIMR Score (1993 to 1995)	-0.285	.	1									
DAC	0.255	-0.253	-0.292	1								
RND	0.376	-0.294	-0.285	0.236	1							
LEV	-0.219	0.033	-0.019	-0.133	-0.245	1						
ln(Size)	0.326	0.057	0.114	-0.108	0.316	-0.058	1					
ln(Anal)	0.180	0.092	0.136	-0.065	0.252	0.025	0.781	1				
EPE	-0.148	-0.235	-0.237	0.237	0.190	0.110	-0.375	-0.028	1			
CV	-0.072	-0.114	-0.136	0.225	-0.035	0.198	-0.293	-0.175	0.279	1		
Ln(INSDR)	0.148	.	-0.007	-0.058	0.191	0.005	0.518	0.451	0.063	-0.127	1	
PINSDR	-0.105	.	-0.124	0.059	-0.142	0.048	-0.355	-0.343	0.012	0.124	-0.123	1
Ln(Vol)	0.215	-0.062	0.052	0.060	0.311	0.066	0.779	0.737	0.053	-0.103	0.447	-0.364

These numbers are calculated as the average of the correlations by year.

Table 4.4: Univariate analysis.

The dependent variable is the adverse selection cost of trading λ, expressed as dollar cost, per \$100 traded.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
S&P T&D	0.0043***						
(S&P T&D)^2	-0.0005**						
AIMR		0.0025***					
(AIMR)^2		-0.000065***					
Ln(Anal)			0.0484***				
(Ln(Anal))^2			-0.0219***				
Abs(CV)				0.0447***			
(Abs(CV))^2				-0.0037**			
Abs(EPE)					0.5409**		
(Abs(EPE))^2					-0.2367**		
MB						0.1183***	
MB^2						-0.0120***	
DAC							0.0227*
DAC^2							-0.0018**
Constant	-0.0005***	0.1150***		0.1228***	0.2143***	-0.1207***	0.2196***
R-square	0.35	0.31		0.13	0.12	0.19	0.05

Note:

1. (***),(**) and (*) indicates significance at the (1%), (5%) and (10%) levels.
2. Models 3 through 7 were run with 9-year dummies corresponding to 1994 to 2002. For sake of brevity, those coefficients not reported. Model 2 contains 2 year dummy (1994,1995).
3. S&P T&D is the S&P 500 transparency and disclosure rank (section 4.4.1.2).
4. AIMR is the annual ranking of corporate disclosure practices published between 1982 and 1996 (section 4.4.1.1).
5. Ln(Anal) is the natural log of the number of analysts providing earnings forecasts for the firm
6. Ln(Anal)^2 is the square of the natural log of the number of analysts providing earnings forecasts for the firm
7. MB is the market to book ratio (section 4.4.2.3).

6. CV is the coefficient of variation of analysts' forecasts (section 4.4.3.3).

7. EPE is the analysts' earnings prediction error (section 4.4.3.2).

8. DAC is discretionary accruals (section 4.4.2.1).

Table 4.5: Multivariate regression analysis.

Equation 4.10 is estimated as a cross-sectional model in every year. The following table presents the time-series averages of the estimated coefficients. The levels of significance represent the result of t-test for mean equal to zero. The adjusted R2 presents the time-series average adjusted R2. The depended variable is the adverse selection cost of trading λ, expressed as a fraction of the bid-ask spread.

	Expected Sign	Model 1	Model 2	Model 3	Model 4	Model 5
MB	+	0.03813***	0.02942***	0.03072***	0.02589	0.01939*
MB^2	-	-0.00444***	-0.00342***	-0.00359***	-0.00879	-0.00238*
lnAnal	+	0.04444***	0.04205***		0.23391*	0.09345*
lnAnal^2	-	-0.01016***	-0.0134***		-0.04297**	-0.01633
CV	+	0.00053*	0.00046*		0.02648	0.00174
CV^2	-	-2.28E-06**	-2.52E-06*		0.01244	-0.02229**
EPE	+	0.02556	0.1236***	0.09824**	0.79032**	
EPE^2	-	-0.02528*	-0.12019**	-0.09829**	-2.85684*	
DAC	+	0.01293	0.01222	0.01567*	0.02466*	0.05233*
DAC^2	-	-0.03102*	-0.00085*	-0.00119*	-0.55361**	-0.30155**
AIMR	+				0.00012**	
AIMR^2	-				-4.59E-06*	
S&P T&D	+					0.00894**
S&P T&D^2	-					-0.00091*
ln(EMP)	+	0.01013**	0.00653*			
(ln(Emp))^2	-	-0.0017*	-0.0012*			
PC_INSDR	+/-	0.00059**			-0.00028	
PC_INSDR^2	+/-	-0.00001			0.00001	
Lev		0.01885*	0.01416	0.01656*	0.11769***	0.05835*
ln(Vol)		-0.03848***	-0.02748***	-0.03258***	-0.06941***	-0.02504***
OI		0.12864***	0.09577***	0.09382***	0.17464***	0.05067
LD		-0.00626	-0.00534	-0.00666	-0.04587*	-0.01841
ID		-0.00781*	-0.00512	-0.00184	-0.02101*	0.02243
(Constant)		0.48423***	0.48661***	0.56181***	0.57155***	0.39546***
Adj R^2		0.451	0.36	0.35	0.684	0.52

Note: (***), (**) and (*) indicates statistical significance at the (1%), (5%) and (10%) levels.

Variable definitions: Table 4.4 notes contains the definitions of MB, ln(Anal), CV, EPE, DAC, AIMR and S&P T&D variables. The additional variables are:

1. PC_INSDR is the per capita insider ownership (percentage of shares held by insiders divided by the number of insiders holding equity ownership).
2. Ln(Emp) is the natural log of the number of employees in the firm
3. LEV is the firm's financial leverage
4. ln(Vol) is the log of mean monthly trading volume in the corresponding year.
5. OrdImb is the monthly order imbalance, expressed as a % of the total order-flow.
6. LD (Loss Dummy) takes the value 1 "earnings before extra-ordinary items" is less than zero, and zero otherwise.
7. ID (Intangible Dummy) takes the value 1 if non-zero R&D expense is reported, zero otherwise.

Table 4.6: **Effect of focus enhancing spin-off.**

The tables below present the results corresponding to 56 focus enhancing spin-off parent firms. The breakpoints (Transparent vs. Opaque) are obtained from Table 4.4; N is the number of firms; DolGH is the dollar adverse selection cost of trading per $100 traded; and GH is the adverse selection cost component expressed as a percentage of the spread.

Panel A: AIMR disclosure ranks (Data 1993, 1994, and 1995)

	Transparent (AIMR rank > 40)				Opaque (AIMR rank > 40)				Total			
Event Month	N	DolGH	GH	Spread	N	Dol GH	GH	Spread	N	Dol GH	GH	Spread
-2	7	0.0861	0.1869	0.1830	12	0.1017	0.2099	0.2204	19	0.0960	0.2014	0.2066
-1	7	0.0845	0.1586	0.1784	12	0.1070	0.2340	0.2398	19	0.0987	0.2062	0.2172
0	7	0.0817	0.1834	0.1839	12	0.1196	0.2200	0.2373	19	0.1056	0.2065	0.2176
1	7	0.0810	0.2047	0.1828	12	0.1292	0.2358	0.2752	19	0.1114	0.2243	0.2412
2	7	0.0777	0.1949	0.1827	12	0.1253	0.2370	0.2308	19	0.1078	0.2215	0.2131
3	7	0.0791	0.2050	0.1816	12	0.1254	0.2452	0.2288	19	0.1083	0.2304	0.2114

Panel B: Analysts' earnings forecasts coefficients of variation (Data 1993 to 2002)

	Transparent (Abs(CV)<=0.18)				Opaque (Abs(CV) > 0.18)				Total			
Event Month	N	DolGH	GH	Spread	N	Dol GH	GH	Spread	N	Dol GH	GH	Spread
-2	14	0.2398	0.2278	0.1671	42	0.1064	0.2544	0.1485	56	0.1398	0.2478	0.1532
-1	14	0.2482	0.2664	0.1711	42	0.0991	0.2459	0.1485	56	0.1364	0.2510	0.1542
0	14	0.1691	0.2584	0.1568	42	0.1380	0.2511	0.1530	56	0.1458	0.2529	0.1540
1	14	0.1761	0.1889	0.1432	42	0.1312	0.2624	0.1542	56	0.1424	0.2440	0.1515
2	14	0.1593	0.2559	0.1632	42	0.1130	0.2264	0.1493	56	0.1246	0.2338	0.1528
3	14	0.1864	0.2447	0.1497	42	0.1265	0.2411	0.1533	56	0.1415	0.2420	0.1524